Evaluating and Improving the WINGS for kids After-School Social and Emotional Learning Program

April 2017

Project Staff

Cara Adams
Julie Blodgett
Alexis Brewer
Laura Brock (PI‑College of Charleston)
Claire Cameron
Amy Cordier
Elizabeth Cottone
Nancy Deutsch
Justin Dormal
Chelsea Duran
David Grissmer (Project CO‑PI)
Pamela Jiranek
Helyn Kim
Shelley Lieberman
Brittany Lorick
Andrew Mashburn (Project CO‑PI)
William Murrah
Julia Thomas
Karen Walker (PI‑Child Trends)
Rachel Warne
Hall West

WINGS Evaluation-Final Report to SIF

Table of Contents

WINGS Evaluation-Final Report to SIF

Executive Summary

i.   The WINGS Program

WINGS for Kids© is a structured after-school social and emotional learning (SEL) program for children attending four low-performing schools in high-risk neighborhoods in Charleston County School District, South Carolina. Over 90% of the students were black, eligible for free lunch, and at high risk for poor academic and behavioral outcomes. WINGS was designed based on research that suggested that effective SEL programs incorporated components that included (1) high participation rates, (2) a multi- year program, (3) a focus on both academic and social/emotional skills, (4) four "SAFE" characteristics (*sequenced, active, focused, and explicit*), and, (5) a focus on five key SEL competencies: *self-awareness, self-management, responsible decision-making, social awareness,* and *relationship skills* (Zins et al., 2004; Payton et al., 2008; Lauer et al. 2006; Greenberg et al. (2003).

During our study, the WINGS program served approximately 24 children in each grade at each school. WINGS afforded opportunities for children to develop SEL skills using a curriculum that was implemented throughout the program's daily activities that included choice time, free play, academic center time, and meals and/or snacks. WINGS was implemented for three hours per day, five days per week during the school year. The multi-year program allows participation from kindergarten to 5th grade.

At each school, the programs are organized in groups or "nests" of 12 students, with two nests per grade. Each nest is assigned a WINGS Leader (WL) who serves as mentor and teacher to his or her nest for the entire year. The five competencies are addressed across 30 Learning Objectives. Each week a new learning objective is emphasized and previously taught objectives are reinforced. Teaching is initially direct, with follow-up modeling, opportunity to practice skills, and coaching applied to real life lessons, also known as "teachable moments." Learning objectives are intentionally embedded into every program activity. Through these activities, the WINGS staff model each learning objective, and reinforce SEL competencies.

The WINGS program *framework* states that at least two years of participation would be required to see significant shifts in SEL competency. The evaluation theory of action predicts that changes in SEL skills will transfer to more positive and less negative relationships and behaviors particularly in the school classroom, but also at home, and have positive long-term impacts on children's academic outcomes. The study also collected an exploratory set of "building block" measures of early cognitive and emotional skills to better understand the underlying developmental mechanisms leading to the outcomes and to help interpret the pattern of outcomes.

ii. Study Objectives

This report provides **interim** results of an experimental evaluation based on a kindergarten lottery for entry to WINGS. Three major data collections provided 38 confirmatory and exploratory outcome measures including a parent survey, a teacher survey and direct child assessments. Child direct measures and parent surveys were collected in the summer or early fall at pre-test, at post-test after one year, and at post-test after two years. Teacher surveys were collected in the fall and spring of kindergarten and first grade.

This report summarizes the impacts for three cohorts of children after one year of participation, and the impacts for two cohorts after two years of participation. The results for one year of kindergarten participation across three cohorts show a strong pattern of null results. These results are consistent with the WINGS theory of action that states that two years of participation is likely needed to produce significant effects. The primary focus of this interim report is on the impact of two years of participation in WINGS for two cohorts of study participants. These results are preliminary until a final evaluation is completed that includes three cohorts of participants. Final data has been collected for the third cohort in November, 2016, and final results will be available by October, 2017.

This interim two year evaluation of WINGS includes 141 kindergarten children in Cohort 1 and 102 children in Cohort 2 who were randomly assigned to WINGS or control conditions. About 58% of the children were assigned to WINGS and 42% to control conditions. This RCT incorporates mixed methods data collected from parents, teachers and WINGS personnel, individual child testing and observational data that enable a more informed understanding of the effects and potential causative mechanisms.

iii. Background: Research Literature

Children from low-income families, in particular, face many challenges and risks related to their social-emotional development that can have negative consequences later on in life (Duncan & Magnuson, 2005). Unfortunately, gaps in social-emotional development between low-income children and their more affluent peers are observed before entering formal schooling, and these gaps persist and increase during the elementary school years and beyond (Alexander, Entwisle, & Kabbani, 2001; Brooks-Gunn, Duncan, & Aber, 1997; Hamre & Pianta, 2001). Without early intervention in social-emotional and behavioral skills, young children are at greater risk for future academic problems, dropping out of school, peer rejection, and antisocial behaviors (Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011; Durlak & Weissberg, 2011; Greenberg et al., 2003). On the other hand, strengthening young children's social-emotional competence may serve as an important protective factor for school and life success, especially if they are exposed to multiple life stressors (Jones, Greenberg, & Crowley, 2015; Webster-Stratton, Reid, & Hammond, 2004). As such, researchers have begun investigating

promising approaches and intervention programs, ranging from in-school curriculums to teacher and parent training programs, that target the promotion of social-emotional competence in children (e.g., Jones & Bouffard, 2012; Morris et al., 2013; Morris et al., 2014; Webster-Stratton,Reid, & Hammond, 2004). There is less work, though, on the impact of social-emotional competence interventions in after-school settings, despite the fact that these competencies can be taught in various ways across many different settings and contexts (CASEL, 2016).

Unfortunately, it is more difficult to conduct research on the effects of socio-emotional and other non-cognitive skills on academic outcomes, than it is to study the direct improvement of academic outcomes during schooling. The lack of routine data collection on socio-emotional skills, the small number of programs developing these skills, the challenges of working within natural settings and the lack of experimental evaluations has resulted in a research literature that is methodologically limited (Durlak, Weissberg, & Pachan, 2010; Lauer et al., 2006).

Research that is **almost all non-experimental** suggests that SEL programming promotes positive youth development across a wide developmental span; in school- based, after-school, and community settings; with students who do and do not have presenting problems; in urban, suburban, and rural areas; among racially and ethnically diverse student bodies; and as implemented by professionals as well as school staff (Payton et al., 2008; SEL Research Group/CASEL, 2010). Studies have suggested students benefit from SEL across a wide range of outcomes, including having higher quality relationships with peers and adults, having fewer problem behaviors at school, using drugs and alcohol less, engaging in risky sexual behavior less, and behaving violently less (Greenberg et al., 2003). Students also have better attitudes about themselves, others, and school, and earn higher grades and test scores (Durlak, Weissberg, et al., 2011; Payton et al., 2008).

In a meta-analysis of 317 almost all non-experimental studies of SEL programs, Payton et al. (2008) suggested that SEL programming was associated with students' gaining an average of 11 to 17 percentile points on achievement tests. Among the 180 studies of programs considered "universal" (not targeted), the authors found a mean effect on academic performance of .28. Effects on other outcomes such as attitudes toward self and others, positive social behavior, conduct problems, and emotional distress were similarly in the .20 range (Payton et al., 2008). Similarly, in another meta-analysis of 213 studies involving more than 270,000 students, Durlak and colleagues (2011) found that, overall, SEL programs both in and out of school were significantly effective (grand study-level mean = 0.30). Specifically, students who participated in evidence-based SEL programs demonstrated enhanced SEL skills, better attitudes about themselves, others, and school, and increased prosocial behaviors, compared to students who did not participate in these programs. Students participating in the SEL programs also had

fewer conduct and internalizing problems, lower levels of emotional distress, increased ability to manage stress and depression, as well as significant gains of 11 percentile points in academic achievement compared to students in the control group. Follow-up data (at least six months later) showed sustained effects in all areas listed above, with effect sizes ranging from 0.11 to 0.32.

Many interventions and programs targeting the promotion of social-emotional competence aim to promote the building blocks that set a strong foundation for social-emotional development in young children (e.g., Morris et al., 2014). Executive function (EF), in particular, has received much attention given its critical role in the development of social-emotional competence (e.g., Bierman, Nix, Greenberg, Blair, & Domitrovich, 2008; Blair & Raver, 2015; Riggs, Jahromi, Razza, Dillworth-Bart, & Mueller, 2006).

Executive function (EF) is a multi-faceted construct that can broadly be defined as the processes of cognitive flexibility, working memory, and inhibitory control that are necessary for purposeful, goal-directed behavior. Studies show that there are persistent and growing poverty-related gaps, not only in achievement (Reardon, 2011), but also in the regulation of attention, emotion, stress response, and executive function (Cicchetti, 2002; Evans, 2003). Evidence from neuroscientific studies suggest that focusing on EF can enhance children's learning and development and can establish positive academic trajectories, particularly for children from low-income families (Blair & Raver, 2015; Evans & Schamberg, 2009; Raver, Blair, Willoughby, & Family Life Project Key Investigators, 2013).

Specific to children's social-emotional competence, EF has been directly implicated in the concurrent and longitudinal development of social-emotional skills (e.g., Riggs, Blair, & Greenberg, 2003). This is not surprising given the many overlaps between the subskills that underlie both EF and social-emotional development. Studies shows that difficulties in EF lead to difficulties in multiple components of social-emotional functioning, including impulsivity, delay of gratification, emotion regulation, problems with attention, behavioral issues, and problem solving (e.g., Cole, Usher, & Cargo, 1993; Hughes, 2002; Jahromi & Stifter, 2008; Kim et al., 2016; Pennington, 2002; Seguin, Boulerice, Harden, Tremblay, & Pihl, 1999). Moreover, the EF components related to planning, inhibiting response, and controlling one's attention may be particularly useful for resisting temptation, regulating frustration and stress (Mischel et al., 1989), and behaving according to social norms (Peake, Hebl, & Mischel, 2002).

iv. Theory of Action

Figure1 depicts the WINGS program's theory of change. The stated objectives of the program are to enhance children's person-centered competencies (self awareness, social awareness, responsible decision-making, self management, and relationship management), and the theory of change follows that (a) children assigned to WINGS will

develop greater person-centered competencies than children who are not assigned to WINGS. The theory of change also proposes that the positive impacts of assignment to WINGS will transfer to the classroom and home setting. Specifically, we hypothesize that compared to children not assigned to WINGS, children assigned to WINGS will develop and display more positive and less negative behavior and relationships with their teachers and classmates in school and at home. The (b) increased positive and decreased negative relationships and behaviors are hypothesized to be (c) mediated through the improvements in children's enhanced person-centered competencies. Finally, assignment to WINGS is proposed to have (d) longer term positive impacts on children's academic school outcomes and socio-emotional behavior. We also collected a broader set of measures for exploratory analysis that focused on measures of early emotional and cognitive skills including executive function that can contribute to identifying possible causative mechanisms that underlie the impacts as well as interpret the pattern of results across outcome measures.

Figure 1 Theory of Change
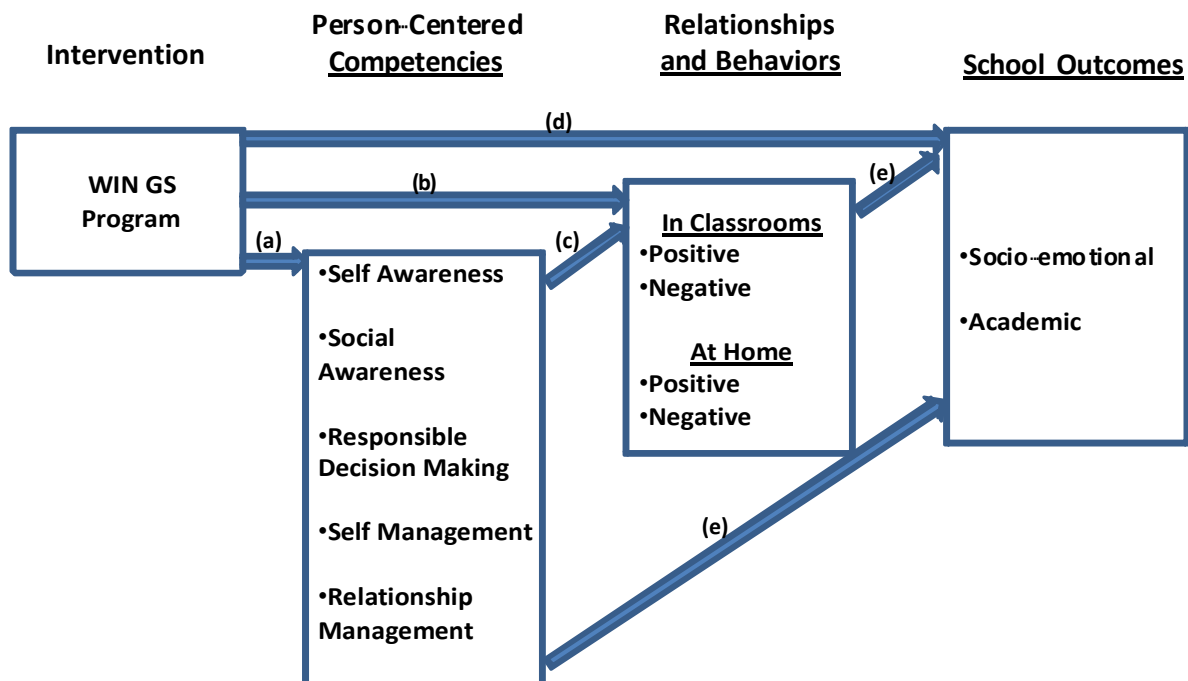
v.   Research Questions

The following research questions are addressed in this study.

Question 1. Does assignment to WINGS have a positive impact on children's person-centered competencies after one year (kindergarten) and two years of WINGS (kindergarten and 1st grade) participation?

Question 2. Does assignment to WINGS have a positive impact on children's relationships and behaviors in the classroom and at home after one year and two years of WINGS participation?

Question 3. Does assignment to WINGS have a positive impact on measures of children's short-term academic skills after one year and two years of WINGS participation?

Question 4. Does the impact of WINGS on children's person-centered competencies, and relationships and behaviors at school and home after one year and two years vary for children with different characteristics?

Question 5. Does the impact of WINGS on children's person-centered competencies, and relationships and behaviors at school and home change from cohort 1 to cohort 2?

Question 6. Does the impact of WINGS on children's person-centered competencies, relationships and behaviors, and school outcomes vary by the level of initial skills?

vi.   Research Design for Current Evaluation

This interim evaluation primarily focuses on the evaluation of two cohorts who receive up to two years of WINGS participation. Cohort 1 had four schools and cohort 2 had only three schools due to the discontinuation of the program at James Simons. James Simons transitioned to a Montessori Magnet school that changed its demographic characteristics, and the WINGS program was discontinued. Specifically, child-level random assignment to WINGS or control was determined within four schools in cohort 1 and three schools in cohort 2 and 3. Because the program serves 12 girls and 12 boys who enter kindergarten each year and conducts social-emotional learning activities separately within each gender "nest", gender will also serve as a randomization block to ensure equal numbers of girls and boys are enrolled in the program. Thus, there will be 14 blocks for the current two cohort study evaluating two years of WINGS and 20 blocks for the three cohort study of one year impacts.

Pre-test data was collected in the summer/fall of kindergarten entrance and post-test data for one year of potential WINGS participation was collected in the spring of kindergarten and the summer/fall following kindergarten. Post-test data for two potential years of WINGS participation was collected in the spring of 1st grade and in the summer/fall following first grade.

Table1 shows that children in the two year, two cohort study were randomized within 14 randomization blocks: two cohorts, at four schools, by gender group (Female/Male), with

one school (James Simons) having only one cohort. Across Cohorts 1 and 2, 101 children (41.7%) were assigned to the control group and 141 (58.3%) were assigned to the treatment group. Table 1 also provides the sample sizes for each randomization block, as well as the probabilities for being assigned to treatment or control within each block. Overall, about 60% of participants were assigned to treatment with 40% to the control group. We assigned more to treatment in order to compensate for expected non-compliance.

Table 1 Randomization Blocks – Frequencies and Probabilities.

Cohort	School	Gender	Random Assignment - Frequencies (n)		Random Assignment - Probabilities	
			Control	Treatment	Control	Treatment
1	Chicora	Female	11	16	40.7%	59.3%
		Male	5	9	35.7%	64.3%
	Memminger	Female	6	9	40.0%	60.0%
		Male	9	13	40.9%	59.1%
	NCES	Female	9	13	40.9%	59.1%
		Male	8	11	42.1%	57.9%
	James Simons	Female	6	7	46.2%	53.8%
		Male	4	4	50.0%	50.0%
2	Chicora	Female	8	12	40.0%	60.0%
		Male	6	8	42.9%	57.1%
	Memminger	Female	7	10	41.2%	58.8%
		Male	6	7	46.2%	53.8%
	NCES	Female	7	11	38.9%	61.1%
		Male	9	11	45.0%	55.0%
TOTAL			101	141	41.7%	58.3%

vii.   Exploratory and Confirmatory Outcome Measures

Assessment tools used in this study included direct child assessments, and measures from teacher and parent reports on classroom and home behavior and relationships. Person-Centered Competencies. Direct assessments were completed in areas that align closely with the constructs of self-awareness, social awareness, responsible decision-making, self-management, and relationship management identified in the theory of change. Parents and teachers reported on the five SEL skills (self-awareness, social awareness, responsible decision-making, self-management, and relationship management) via the *Devereux Student Strengths Assessment* (DESSA; Lebuffe, Shapiro, & Naglieri, 2008).

Teacher-report measures of children's relationships and classroom behaviors included the *Student-Teacher Relationship Scale* (STRS; Pianta, 2001), which measures the quality of the teacher's relationship with individual children, and the *Social Skills Improvement System* (SSIS; Gresham & Elliott, 2008), which is a measure of an individual child's relationships and social behaviors in the classroom. Children's relationships and behaviors at home were assessed during parent/guardian interviews using parent versions of the *Social Skills Improvement System* (SSIS; Gresham & Elliott, 2008) and the *Child-Parent Relationship Scale* (CPRS; Pianta, 1992).

School Outcomes. Direct assessments of academic outcomes were completed using the *Woodcock-Johnson-III Tests of Achievement* (WJ-III; Woodcock; McGrew, & Mather, 2001), which evaluates reading skills (Sound Awareness and Letter-Word Identification subtests), mathematics skills (Applied Problems and Quantitative Concepts subtests), and general knowledge (Academic Knowledge).

Building Block Skills included measures of executive function: *Head-Toes-Knees-Shoulders Task* (HTKS, Ponitz, McClelland, et al., 2008), *Emotion Matching Task* (EMT; Morgan, Izard, & King, 2009), *Assessment of Children's Knowledge Task* (ACES; Mavroveli et al., 2009), *Theory of Mind* (NEPSY II; Korkman, Kirk & Kemp, 2007a, 2007b), *Delay of Gratification Task*, (Mischel, Shoda, & Rodriguez, 1989), and Visual-Motor Integration (VMI; Beery VMI; Beery, 2010; Beery, Buktenica & Beery, 1997).

viii.   Response Rates

Table 2 shows the response rates for each type of assessment at each time point by cohort. Due in large part of the mobility of study families, the prevalence of missing data increased throughout the course of this study, with more missing data occurring for each cohort in later time points than in early time points. For instance, Cohort 1 response rates at baseline (Summer/Fall of Kindergarten) were above 90% for all types of measures, but were less than 80% at the end of 1st Grade and beginning of 2nd Grade.

Table 2. Percent Data Collected (Response Rates) by Time Point, Measure Type, and Cohort.

Time Point	Type of Measure	Cohort 1	Cohort 2
Summer/Fall of Kindergarten	Direct Measures	99%	98%
	Parent Reports	92%	97%
	Teacher Reports	93%	96%
Spring of Kindergarten	Teacher Reports	96%	97%
Summer/Fall of First Grade	Direct Measures	95%	84%
	Parent Reports	89%	75%
	Teacher Reports	86%	53%
Spring of First Grade	Teacher Reports	73%	60%
Summer/Fall of Second Grade	Direct Measures	77%	63%
	Parent Reports	73%	63%

Studies of children that occur outside of the regular attendance during schools days encounter challenges in tracking and collecting data from parents and testing children. Low-income, urban families are a particular issue due to their frequent relocation of households and changing schools for their children. Table 3 shows the percentage of children that were enrolled in a non-study school by the summer of 2015, approximately three years after the start of the study for Cohort 1 and two years after the start of the study for Cohort 2. This data suggests an annual migration of 20% of children relocating to non-study schools during the study. Study children that originally attended four study schools are currently dispersed across at least 52 different schools, only ten of which are outside of South Carolina. Part of the cause of the relocation can be changes in jobs or income that demand a move.

Table 3 Percent of Children Relocated by Cohort and School.

School attended at start of study	Percent relocated to non-study schools as of Summer, 2015	
	Cohort 1	Cohort 2

Chicora	39%	29%
Memminger	68%	43%
NCES	68%	34%
JSE	43%	N/A
Total	56%	35%

ix. Attrition and Noncompliance and its Significance

Attrition. Differences in response or attrition rates between test and control groups can be problematic and pose a potential source of bias in effect estimates depending on whether the characteristics of those students with data are similar to those without data. Table 5.3 provides response rates by major data collections (parent, teacher, and child testing) by treatment and control group for the one year of participation sample and the two year of participation sample. The table also shows the tests for the statistical significance of the differences. Only 1 of 12 comparisons between treatment and control groups shows a significant difference. Only two of twelve comparisons do not meet conservative WWC attrition and differential attrition standards, and all comparisons meet the liberal WWC standards.

This suggests that random assignment was balanced on demographic characteristics. The only differences between treatment and control groups were in number of children in the home and occurrence of stressful life events in the year prior to the start of the study. We have included covariates to adjust for these differences.

Sample Non-Compliance. In contrast to the term "study attrition," which refers to individual children or families not being available to provide data for the *study,* the term non-compliance refers to whether the participants complied with their treatment and control assignment. Participants who won the lottery are non-compliers if they do not meet the standards for completing either one or two years of WINGS. Participants who lost the lottery are non-compliers if they actually attend the WINGS program and meet the standards for one and two years of attendance. These participants are termed crossovers. Non-compliance can introduce bias if the non-compliers' characteristics are not similar to compliers for both treatment and control groups. Non-compliance also is problematic because it reduces the statistical power of the sample.

According to the WINGS logic model, attending WINGS for two years is thought to be necessary before seeing positive impacts. In conjunction with WINGS personnel, we also established minimum attendance criteria for each year based on actual attendance data. We set the criteria of at least 100 days of attendance in kindergarten and first grade to qualify as having "received treatment." Based on these criteria, Tables 5.6 and

5.7 provide the consort data for the level of compliance and non-compliance for Cohort 1 and Cohort 2 respectively.

Tables 4 and 5 show that only 30 of 82 (37%) participants assigned to treatment in Cohort 1 met the attendance criteria in both kindergarten and 1st grade, while the compliance rate for treatment in Cohort 2 was 42%. The compliance rates for one year of attendance were much higher with 68% (Cohort 1) and 61% (Cohort 2) of treatment children complying. The compliance rates for control children were very high with 86% (Cohort 1) and 91% (Cohort 2) of children not receiving treatment.

Table 4 Cohort 1 Two- Year Consort Data

Randomized Children:	140							
Treatment Condition:	Treatment (n=82)				Control (n=58)			
Attended at least 100 days in Kindergarten?	Yes		No		Yes		No	
	56		26		1		57	
% of condition group	68%		32%		2%		98%	
Attended at least 100 days in First Grade?	Yes	No	Yes	No	Yes	No	Yes	No
	30	26	8	18	0	1	7	50
% of condition group	37%	32%	10%	22%	0%	2%	12%	86%

Table 5 Cohort 2 Two- Year Consort Data

Randomized Children:	102							
Treatment Condition:	Treatment (n=59)				Control (n=43)			
Attended at least 100 days in Kindergarten?	Yes		No		Yes		No	
	36		23		1		42	
% of condition group	61%		39%		2%		98%	
Attended at least 100 days in First Grade?	Yes	No	Yes	No	Yes	No	Yes	No
	25	11	5	18	1	0	3	39
% of condition group	42%	19%	8%	31%	2%	0%	7%	91%

The study maintained records for each child that drew from parent conversations and WINGS personnel and recorded the reasons for treatment children withdrawing from the program. Table 6 shows these reasons for non-compliance. About 60% of non-compliance for treatment children was due to a relocation and attendance at another

school without the WINGS program. The WINGS program was available only in the four study schools, so almost all relocation was to district schools without WINGS. The second most important reason for non-compliance (23%) was removal by the parent without relocation. Removal by the parent could occur for a wide number of reasons and might reflect the increased stress on the child and parent from the longer day at school, and the parent preference to have the child return home after school. Parents respond very positively when asked about the WINGS program, and there is little evidence that dissatisfaction with WINGS is a significant cause of removal.

Table 6 Percent of Children Relocated by Cohort and School.

School attended at start of study	Percent relocated to non-study schools as of Summer, 2015	
	Cohort 1	Cohort 2
Chicora	39%	29%
Memminger	68%	43%
NCES	68%	34%
JSE	43%	N/A
Total	56%	35%

x.   Estimation Methodology

We estimated ITT and TOT impacts of WINGS on all 38 measures for both one year of participation using three cohorts of data and two years of participation using two cohorts of data. The ITT analysis provides estimates of the impact on all children assigned to WINGS, regardless of whether they attended, while the TOT analyses provides estimates of the impact for children who actually attended- although the estimates remain experimental only if compliers and non-compliers are similar. The TOT effects show much larger effects than ITT effects because, for instance, only approximately four out of ten treatment children completed two years of WINGS.

We estimate ITT effects using the standard estimation procedures incorporating a treatment variable, pre-tests and covariates. Child gender, cohort, and school were each included in the covariates list since these were variables on which random assignment was blocked. Including child age improves precision by removing variance in child outcomes due to age/maturation. Inclusion of child pre-test scores allows for interpretation of impact estimates as being on *development* or *improvement* in child outcomes. Mother's education and age at child's birth were included as they were related to many child outcomes, and therefore can also improve precision of estimates.

Number of children in the home and parent stress was not balanced between the assigned treatment and control groups; therefore, these variables are included as covariate to achieve balance between randomly assigned groups. We also add covariates to account for differences in the time between pre- and post-test and whether the test occurred in the summer or during the school year.

We estimate the TOT effect of attending WINGS using an instrumental variables (IV) approach to account for "no shows," or those who choose not to attend after being offered admission, and "crossovers," or those who attend WINGS after not being offered admission (Bloom, 1984). In general, IV requires at least one instrument that is correlated with the endogenous choice (whether or not to attend) but is otherwise uncorrelated with the outcome (achievement). If students with an offer are more likely to attend, it satisfies the first requirement of an instrument. Because offers were random, we assume the lottery outcome is otherwise uncorrelated with outcomes.

Experiments that have non-compliance, missing data, differences in measure reliability, multiple outcome measures, and can encounter improbable random draws require a more robust analysis that involves doing extensive sensitivity analysis of the results to a variety of assumptions.In this report, we provide ITT and TOT estimates for two distinct samples: the sample that includes only cases with complete data (list-wise deletion) and the full sample with imputations for all missing data. We also provide results for three different specifications. The first specification contains the treatment variable with no covariates, while the second specification adds the pre-test as a covariate. The third specification adds a list of additional covariates to control for variations in random assignment and variations In general, the pattern of results did not show much sensitivity to inclusion of missing data or to alternate specifications.

The size of our randomized sample for two cohorts is 242 children, and we have complete data (listwise deletion) for all outcomes and covariates for154 children. Results using listwise deletion are viable only if attrition is random. One characteristic of the missing data is that it is seldom at the item level alone, but rather is due to an entire missing teacher survey, parent survey, or lack of all child assessments at least at one point in time. We have relatively complete data from all data sources at the pre-test, but due to migration, parent and teacher surveys and child assessments became increasingly more difficult to gather at later time points.

The extent of missing data by source of data meets, for the most part, conservative WWC standards, and all sources of attrition meet liberal WWC standards. However, the extent of missing data was large enough, and when combined with the levels of non-compliance and associated loss of statistical power, the listwise results alone would leave important stability issues. Thus, more complex methods needed to be incorporated into our analyses.

The inclusion of missing data into both the ITT and TOT estimates posed a challenge because the most frequently used procedure for incorporating missing data is full maximum likelihood estimates (FIML). However, FIML estimation cannot be incorporated into TOT estimates because of the two-stage procedure. Since TOT estimates are a critical component in interpreting the impacts of WINGS, we utilized a second procedure for handling missing data - Multiple Imputation (MI). MI can be utilized to produce both ITT and TOT estimates using the same assumptions and framework. MI is a modern and powerful missing data handling approach, and, like FIML, is also considered state-of-the-art (Enders, 2010). In contrast to FIML, which serves to provide the most probable parameter estimates that fit a given data set, the MI approach begins by filling in – i.e., imputing – probable values in place of missing data points. This process makes use of information from predictor variables as well as auxiliary variables (which do not appear in the model but are related to missingness in a particular variable and/or the variable itself), and also explicitly preserves uncertainty of imputed estimates. The imputation algorithm is repeated many times to create multiple imputed data sets, across which analyses can be run and then pooled to provide reliable parameter estimates.

xi.   Results

The ITT results for one year of participation showed a overall pattern of null results with only a few reaching the $p < .20$ level of significance and none reaching the $p < .05$ level of significance. A supplementary spreadsheet with the results across measures is available upon request. In general, the pattern of results did not show much sensitivity to inclusion of missing data or to alternate specifications.

We report ITT and TOT impacts for two years of participation in WINGS for all 38 measures using the full sample with multiple imputation of all missing data and the specification incorporating all covariates. We report three levels of "statistical significance" at the $p < .05$, $p < .10$, and $p < .20$ level because lower levels of significance are important for providing guidance on improving the program, for interpreting patterns of results across measures, and possibly foreshadowing future effects with larger samples. We evaluate the magnitude of effect sizes by the conventional scale of small effect (~.25 SD), moderate (~.50 SD), and large (~.75 SD).

<u>Teacher Assessment of Positive Behavior</u>. Eight of the nine measures show positive effects with effect sizes ranging from .08 to .20. Self management (effect size = .20, $p < .20$) and decision-making (effect size = .18, $p < .20$) show the strongest effects. Figure 7.2 (and Table 7.2) shows the TOT effects where the positive results for eight of nine measures range from .21 to .52. Self management (effect size = .52, $p < .20$) and decision-making (effect size = .42, $p < .20$) show the strongest effects in the moderate range.

Teacher Assessment of Negative Behavior. All six measures show reduced negative behaviors with effect sizes ranging from -.12 to -.30. Four of the measures show significance of $p < .10$. These include hyperactivity (effect size = .31, $p < .10$), problem behaviors (effect size = .27, $p < .10$), bullying (effect size = .26, $p < .10$), and teacher conflict (effect size = .24, $p < .10$). Externalizing behavior (effect size = .23, $p < .20$) also shows reduced incidence. The TOT estimates are in the moderate to large range.   These include hyperactivity (effect size = .77, $p < .10$), problem behaviors (effect size = .69, $p < .10$), bullying (effect size = .68, $p < .10$), teacher conflict (effect size = .63, $p < .10$) and externalizing behavior (effect size = .59, $p < .20$).

Parent Assessment of Positive Behavior The estimates for parent assessment for six positive behaviors suggest null results. The ITT effects range from +.19 to -.18, but none reach the $p < .20$ level of significance. In contrast to teachers, parents report no significant change in positive behaviors for their children.

Parent Assessment of Negative Behavior The ITT and TOT effects for parent reports on negative behavior show reduced or no change in negative behavior on all six ITT measures, but none reach the level of $p < .20$.

Building Block Skills from Direct Child Measures The building block exploratory measures show strong results for two skills: HTKS (effect size = .26, $p < .10$) is a measure of executive function and DAS naming vocabulary (effect size = .26, $p < .05$) is a measure of verbal and comprehension skills. The corresponding TOT effects are in the moderate to large range: HTKS (effect size = .67, $p < .10$) and DAS naming vocabulary (effect size = .57, $p < .05$). None of the remaining six measures reach significance of $p < .20$.

Academic Skills from Direct Child Measures None of the academic measures show positive effects with significance $p < .20$. However, WJ Applied Problems shows a negative ITT effect (effect size = .21, $p < .20$).

Interaction Effects The interactions by gender, school and initial skill level showed no consistent and significant patterns of effects across measures. However, a consistent cohort effect emerged across most measures that suggested WINGS impacts were larger in Cohort 2 than Cohort 1. Our mixed methods data also suggest that the quality of the WINGS program may have been adversely impacted for Cohort 1 both by geographical relocation of one of the schools and from WINGS program expansion to Atlanta and relocation of staff. In addition our data suggests that parent stress levels were much higher in cohort 1 than cohort 2 that may be linked to the school relocation and associated busing of child over longer distances.

xii.   Discussion of Results

The pattern and level of significance of the evaluation results for WINGS suggest that children attending WINGS in both kindergarten and 1st grade have small to moderate impacts on children's positive classroom behavior and moderate to large impacts on reducing negative classroom behaviors. Many of these are in the $p < .20$ and $p < .10$ range, but none reach the $p < .05$ level. We also find that moderate to large gains on two important foundational cognitive skills of executive function and naming vocabulary at the $p < .10$ and $p < .05$ level, but the remaining six building block measures show no consistent effects reaching the $p < .20$ level.

These classroom behavioral effects might be linked to the effects for executive function and naming vocabulary. Execution function is predictive of less impulsive and more focused classroom behavior as well as improved social interactions. Improved social interactions might be linked to both less negative classroom behavior and improved vocabulary. The combination of improvement in two important foundational skills coupled with improved classroom behavior provides additional credibility to the results. Over 80% of measured interventions funded by IES find no significant effects, and those with effects are usually in the small to moderate range. Effects of this magnitude are rarely found in educational and social interventions. These effects are similar to or larger than effects measured in experimental evaluations of school-based socio-emotional programs. For instance, Blair and Raver, 2014 found effects of .3 to .8 for their low income sample on measures of EF and vocabulary. .

The absence of similar effects from parents and teachers suggest two hypotheses. The first hypothesis is that the child's classroom behavior has improved, but behavior at home has not improved. The second hypothesis is that parent ratings are less objective biased by parental stress compared to teacher ratings. Attendance at WINGS may make home behavior more challenging for children and parents. Our data shows that over 65% of control children who arrive home after school have parent/adult supervision with others attending after-school care programs. Home behavior may be different for children attending WINGS who arrive home much later after a long day at school and WINGS compared to children who arrive home after school and spend time with adults. We also find that treatment mothers are more stressed than control mothers. This increased stress may result from the challenges of having a child attend WINGS and the associated challenges of late home arrival of a tired child. The classroom environment is also much more similar to and places similar demands on children as the WINGS environment, so transfer of new learned behavior during WINGS may be easier to the classroom than to the home environment.

Parent ratings of their own children can also be biased by the lack of a peer control group for comparisons that teachers have. Classroom behavior during the day for an

entire school year provides an environment where a child's behavior can be compared to peers. Our data also suggest that parent ratings are strongly influenced by parent characteristics including the stress level of the parent. Parents of WINGS children, other things equal, are more stressed, and stressed parents rate children's behavior lower.

The results also suggest that only attending WINGS in kindergarten does not produce these positive impacts found after two years. This pattern of impacts has three possible interpretations. One interpretation is that WINGS effects occur only if children attend for two years- a single year of attendance provides insufficient dosage for significant impacts. A second interpretation is that the WINGS program is not effective for kindergarten students, but the program is more effective for 1st grade students. A third interpretation is that the current two-year results partly reflect both longer term effects from kindergarten and short-term impacts from 1st grade. If effects are delayed, then effects will increase in the longer term.

Each of these interpretations would suggest that WINGS impacts would increase with more dosage, multiple years of attendance and/or with older children and/or if long-term impacts are measured. The WINGS program serves children through 5th grade, and the current evaluation does not include older children or children who receive more than two years of dosage or measure long-term results. Our evidence would suggest that effects may grow with more dosage, for older children, and may have delayed effects. Thus, the current evaluation that includes only younger children with limited dosage and measures short-term impacts may underestimate the full impact of the program.

WINGS is currently configured with only a few schools in a school district offering WINGS. For children who remain in schools that have the WINGS program, it is possible to attend up to six years since the program serves children through 5th grade. However, our data suggests two factors that make it difficult for children to receive higher levels of dosage. These factors are the high rate of migration of low-income families within school districts to schools not having a WINGS program, and the turbulence and stress present in these families from more frequent changes in jobs, income, health, and relationships that prevent regular attendance at WINGS. For instance, only four in ten children who were given the opportunity to start WINGS at kindergarten attained two years of treatment. Parents of WINGS children express strong approval and support for the program, but regular attendance requires remaining near schools offering WINGS and a level of family stability and commitment that some families find difficult to attain.

One direction for increasing the number of children who can have the opportunity to receive two or more years of WINGS from K-5 is to undertake a district-level demonstration project making WINGS available in all schools in an urban school district. Doing so might substantially increase the opportunity for children to receive more than

two years of dosage, and increase dosage for older students. Such a project would allow students who move within a school district to maintain WINGS attendance, and allow students who drop out of WINGS for a year or more to return and receive additional dosage. There are school districts throughout the nation who are experimenting with extended-day programs, but often utilize the extended time for more direct instruction. However, there is strong experimental evidence that more direct instruction in reading and math in extended-day programs does not lift achievement (Black, et al, 2009). However, the current evaluation results for WINGS appear strong enough to support a larger demonstration project in a school district that would allow experimental measurements of the long term impacts for older children and for children who receive more than two years of dosage.

xiii.   Placing the Results in a Wider Context

Research on children's short- and long-term outcomes and the experimental evaluation of school-based programs and out-of-school programs to improve outcomes are undergoing rapid evolution and development. Historically, most of this research and evaluation was focused on the impact of schools and school-based interventions that have reading and/or math achievement as primary outcomes. School-based interventions using achievement as outcomes provide an advantage in evaluation because the measures are highly reliable and replicable, non-compliance and missing data are usually not problems, statistical power is high from large samples, and outcomes measures are few. In this evaluation environment, the standard ITT and TOT methods of evaluation using statistical significance levels of $p < .05$ are warranted. The best studies also incorporate multiple methods so that causative mechanisms can potentially be identified, and assessment can be done about whether some children benefit more than others and whether there are ways of improving the intervention through improved fidelity.

However, research is increasingly suggesting that children's long-term outcomes are predicted as much or more by developmental skills learned outside of direct school instruction than by skills learned in school direct instruction. These skills include self regulation, executive function, socio-emotional, visuo-spatial, and early comprehension (Grissmer et al, 2010). These early developmental skills have less reliable measures, require a larger set of outcome measures to capture their effects, and need more complex interventions for their improvement because they are learned largely outside schools. These interventions focused on activities outside of schools make large samples less accessible and have much higher levels of non-compliance and missing data that make adequate statistical power more difficult.

The measurement of these early developmental skills is a work in progress, and the quality and reliability of these measurements have not approached those of the most commonly used measures associated with achievement. Part of the problem is that any measure- achievement or otherwise- made with younger children have less quality and reproducibility. We should not expect at this stage of development to have the same kind of results that would be had using achievement measures because achievement measures are narrower, better developed – whereas these skills are displayed in a much wider set of behaviors, more difficult to measure, and less reproducible. These measures will improve over time and allow better measures- but at this stage of development, the criteria should not be the most stringent levels of statistical significance, i.e., $p < .05$. In the long run when measures and interventions are improved and samples have more power, imposing a statistical significance standard seems reasonable, but the purpose of an evaluation during the evolutionary period of improving

measures and programs with weaker samples should incorporate a different set of objectives.

Evaluation methodology during the evolving period when measures and interventions are improving should incorporate the following elements:

- The standard RCT ITT and TOT analysis should be used that include methods of incorporating missing data with particular emphasis on the TOT effect sizes.
- Lower levels of statistical significance should be reported on measures
- Assessing ways of improving measures should be an important objective including incorporating a much wider range of exploratory outcome measures than is typical in RCTs.
- Assessing ways of improving the effectiveness of the intervention is also an important objective to undertake in addition to standard evaluation analysis.
- Interpretations should focus on the internal consistency and predictability from the theory of action of the broad patterns of results across measures rather than consideration of statistics across individual measures.
- No consensus exists about statistical adjustments for multiple comparisons when virtually all outcome measures are correlated. Any consideration of adjustments should recognize that while such adjustments decrease the chances of a false positive (an ineffective program will be labeled as effective), they increase the chances of a false negative (an effective program will be labeled ineffective). In the evolutionary research stages of developing improved measures and programs, searching for patterns linked to effective programs may be more important than guarding against misallocating resources to an ineffective program. Multiple adjustment comparisons do not generally shift the pattern of results that is the major focus of interpreting results during the early evolutionary stages. So they provide little useful additional information. As measures and programs improve and decisions are made involving significant resource allocations, tighter standards become more important.

xiv. Assessing the Current Level of Evidence

A major goal of this project was to raise the *level of evidence for the effectiveness of WINGS from preliminary to moderate.* The preliminary evidence was based on several earlier studies. The first source is a series of master's thesis and unpublished studies that suggest that WINGS participants have better grades, state test scores, school attendance, classroom behavior, self-esteem and higher high school graduation rates compared to student not in WINGS (McGinley et al, 2005; Ivcevic & Brackett, 2005; Ivcevic, Rivers, & Brackett, 2004; Abry, Brock, & Rimm-Kaufman, 2010). The second source is based on two extensive bodies of research about the characteristics of high impact SEL programs (e.g., Durlak, Weissberg, et al., 2010) and after-school programs (Durlak, Mahoney, et al., 2010; McComb & Scott-Little, 2003; Kane, 2004). WINGS includes program components that are characteristic of high

impact programs including high participation rates, multi-year, focus on both academic and social/emotional skills, aligned with SAFE (*sequenced, active, focused, and explicit*) criteria and a framework focused on five key SEL competencies (Zins et al., 2004; Payton et al., 2008; Lauer et al. 2006; Greenberg et al. (2003). However, findings have been inconsistent across studies and programs, and research methods rarely are experimental or quasi-experimental (Kane, 2004; Lauer et al., 2006; Riggs & Greenberg, 2004).

This evaluation of WINGS was designed to provide a *moderate level of evidence on the impact of the WINGS program* with funding from the Institute for Education Science (IES) and the Social Innovation Fund (SIF). The study was designed to utilize a randomized lottery before kindergarten entrance to offer parents the opportunity for participation of their children in WINGS. Four cohorts of children entering kindergarten from 2012-2015 in four (2012) or three (2013-2015) Charleston, SC schools would participate in the study. The total lottery sample was projected to be 440 students. These students will be annually tested until July, 2016, and the sample will include students having four, three, two and one year of WINGS participation, thereby allowing impact estimation for multi-year participation. The lottery design combined with the high statistical power of the study resulting from the large sample should insure high internal validity of the results. The study participants are at high risk for poor academic and behavioral outcomes, and study results can be generalized to populations with similar characteristics.

The original design would have addressed most threats to internal validity (e.g., selection bias, regression artifacts, and ambiguous temporal precedence) would have been addressed through random assignment of each child to WINGS or control conditions. It is expected that following random assignment, children assigned to the WINGS and control conditions will be, on average, equal on all measured characteristics prior to treatment. Thus, any differences in subsequent child outcomes are likely attributable to the one key difference between them–assignment to WINGS or the control group. However, the actual study introduces three important threats to internal validity– overall attrition, differential attrition and non-compliance that introduce a potential for bias and lowered statistical power.

The critical question is the extent to which the higher rates of attrition and much lower rates of compliance due primarily to family migration affected the impacts and their statistical significance. Since family migration rates and attrition were similar for test and control groups, the results would not be expected to be biased. However, higher levels of attrition in both test and control groups and much higher non-compliance in the treatment group would be expected to significantly affect statistical power for ITT estimates. Our estimates and the empirical results suggest that while the originally designed study had minimum detectable effects (MDE) at the .05 level between .15 and

.20, the actual sample after attrition and noncompliance had MDE between .25-.30.

Two reasonable interpretations of the results are that a stronger sample from lower attrition and noncompliance would have produced ITT effects that showed (1) statistically significant results at the .05 level for most of the confirmatory teacher measures as well as executive function and naming vocabulary, and (2) ITT effects would approach the higher level of the TOT estimates. If so, these results would provide much stronger than moderate evidence for effectiveness.

An important consideration in assessing the level of evidence is to identify the specific policy decision that the evidence is meant to support. If the results of this study were used to determine whether the WINGS program would be implemented nationally at substantial costs, then the level of evidence should be stronger than if the results are used to determine whether evidence is sufficient to expand the program within a school district and to provide time and resources to further evaluate and improve the program (as called for in this report). In our view, the current two year results provide moderate evidence for the effectiveness of the WINGS program to continue expansion and to implement improvements that will strengthen the program- particularly if implemented in a district-wide context.

The study has identified several ways to improve the program. One critical factor for increasing its effectiveness is to insure that more children have the opportunity to attend the program for multiple years, and receive higher dosage within each year. Our results for one and two years of attendance would suggest that attendance for two years provides much stronger results than one year and results may be stronger for older children. Family migration within school districts is a major factor in preventing children from attending for multiple years, and receiving more dosage per year. A district-wide WINGS program could provide a much better opportunity to evaluate the program because attrition and noncompliance would be much lower. Our results provide the moderate level of evidence to support this policy direction.

This moderate level of evidence is supported by the following factors:

- The design incorporated a lottery based evaluation incorporating a pre-test
- There was only 2 out of 20 tested variables that showed statistically significant differences at the .10 level at pre-test, and our estimation included covariates for these variables. The attrition levels met WWC liberal levels for all measures.
- The pattern of results consistently supported the theory of action for classroom rated improvement in 8 socio-emotional skills and 6 measures of reduction in negative behavior. In addition, two of the underlying developmental skills hypothesized to contribute to building socio-emotional skills (executive function and naming vocabulary) were significant at the .10 or .05 level.

- The significance level for most of these measures would have plausibly moved into the .10 and .05 level absence the higher than expected level of attrition and noncompliance, and the effect sizes would have approached the much higher levels of the TOT estimates.
- The lack of results linked to the parent measures can plausibly be linked to several factors
  - the lack of objectivity of parent observations
  - the lack of consistent experience with a wider set of peers to assess their child's improvement
  - the higher level of stress measured for WINGS parents may have affected their ratings
  - Home behavior may not have improved because of the longer school day and more stressed parent
  - The WINGS program is better designed to affect classroom behavior than the more challenging home behavior.

Applying adjustments for multiple comparisons would reduce the significance levels for all measures, but retain the pattern of results. However, the issue of whether to apply corrections for multiple comparisons and how to interpret the results is more complicated. Multiple comparisons adjustments have two effects. They reduce the chances of judging a program effective when it is ineffective, but they also increase the chances of judging an effective program as ineffective. The key question is what are the risk and costs of making each type of error. If the results of this study were used to determine whether the WINGS program would be implemented nationally at substantial costs, then applying corrections for multiple adjustments seems appropriate because of the large costs involved if the program turns out to be ineffective. However, if the results of the study are used to determine whether evidence is sufficient to expand the program within a school district and to provide time and resources to further evaluate and improve the program (as called for in our report), then it may be more important to view unadjusted estimates to increase the chances that effective programs are not discarded.

Chapter 1　　Introduction

WINGS FOR KIDS (WINGS) offered an after-school social and emotional learning (SEL) program for K-5th grade children attending four elementary schools located in high-risk neighborhoods in North Charleston, South Carolina. Over 90% of the students were black, eligible for free lunch, and at high risk for poor academic and behavioral outcomes. During our study, the WINGS program served approximately 24 children in each grade at each school. WINGS afforded opportunities for children to develop SEL skills using a curriculum that was implemented throughout the program's daily activities that included choice time, free play, academic center time, and meals and/or snacks. WINGS was implemented for three hours per day, five days per week during the school year.

This report provides experimental impact results based on a kindergarten lottery for entry to WINGS for two cohorts of kindergarten applicants. This randomized control trial (RCT) tracked each cohort from kindergarten entrance through 2nd grade entrance in order to allow children two years of possible WINGS participation. The three major data collections that provided the 38 confirmatory and exploratory outcome measures were a parent survey, a teacher survey and direct child assessments. Direct child direct measures and parent surveys were collected in the summer or early fall at pre-test, at post-test after one year, and at post-test after two years. Teacher surveys were collected in the fall and spring of kindergarten and first grade.

The study includes 141 kindergarten children in Cohort 1 and 102 children in Cohort 2 who were randomly assigned to WINGS or control conditions. About 58% of the children were assigned to WINGS and 42% to control conditions. This RCT incorporates mixed methods data collected from parents, teachers and WINGS personnel, individual child testing and observational data that enables a more informed understanding of the effects and potential causative mechanisms.

The specific objectives of the WINGS program are to improve children's SEL competencies in five areas: *self-awareness, self-management, responsible decision-making, social awareness,* and *relationship skills.* The WINGS program *framework* states that at least two years of participation would be required to see significant shifts in SEL competency. The evaluation theory of action predicts that changes in SEL skills will transfer to more positive and less negative relationships and behaviors in the school classroom and at home, and have positive long-term impacts on children's academic outcomes. In addition, the study also collected an exploratory set of "building block" measures of early cognitive and emotional skills to better understand the underlying developmental mechanisms leading to the outcomes and to help interpret the pattern of outcomes.

We summarize our project in the following eight chapters. Chapter 2 provides the description of the WINGS program. Chapter 3 provides a literature review, the theory of action, research questions, and the research design. Chapter 4 describes the major data collections, their timelines, and response rates. Chapter 5 provides the characteristics of the sample, the randomization process and its outcome and an analysis of study attrition and non-compliance. Chapter 6 provides a description of the outcome measures and the analytical methods used to assess the program impacts including the intent-to-treat (ITT) and treatment of the treated (TOT) estimation and handling of missing data. Chapter 7 provides the results. Chapter 8 provides a discussion of the results, their implications, and future directions for research and the WINGS program. Chapter 9 provides an assessment of the fidelity of implementation, the external and internal environmental factors that might affect impact results, and suggested ways of improving the WINGS program.

Chapter 2   Description of the Program

2.1 Program activities and components

WINGS for Kids© is a structured after-school social and emotional learning (SEL) program for children attending low-performing schools in high-risk neighborhoods in Charleston County School District, South Carolina. At each school, the programs are organized in groups or "nests" of 12 students, with two nests per grade. Each nest is assigned a WINGS Leader (WL) who serves as mentor and teacher to his or her nest for the entire year. The WINGS program–delivered three hours per day, five days per week during the school year–affords opportunities for children to develop SEL skills using a curriculum that is implemented throughout the program's daily activities that include choice time, free play, academic center time, and meals and/or snacks. The multi-year program allows participation from kindergarten to 5th grade. The goal of the WINGS program is to improve children's SEL competencies in five areas: self-awareness, self-management, responsible decision-making, social awareness, and relationship skills. Improvements in these five competencies, in turn, are intended to have a positive impact on children's relationships and behaviors in classrooms and at home, and their social and academic performance in school.

WINGS was designed based on research that suggested that effective SEL programs incorporated components that included (1) high participation rates, (2) a multi-year program, (3) a focus on both academic and social/emotional skills, (4) four "SAFE" characteristics (*sequenced, active, focused, and explicit*), and, (5) a focus on five key SEL competencies: *self-awareness, self-management, responsible decision-making, social awareness,* and *relationship skills* (Zins et al., 2004; Payton et al., 2008; Lauer et al. 2006; Greenberg et al. (2003). The five competencies are addressed across 30 Learning Objectives (see Appendix A). Each week a new learning objective is emphasized and previously taught objectives are reinforced. Teaching is initially direct, with follow-up modeling, opportunity to practice skills, and coaching applied to real life lessons, also known as "teachable moments." Learning objectives are intentionally embedded into every program activity. Through these activities, the WINGS staff model, teach, and reinforce SEL competencies.

The program components (see Appendix B) include:

1) Community Unity- a 30-minute period with the nest to talk about and practice learning objectives and recite the WINGS Creed
2) Choice Time- a twice weekly activity that children choose for a semester, which includes art, music, sports and various thematic units.
3) Academic Time-a daily homework session with adult assistance.

4) WINGSWorks- activities based on helping others through weekly community service projects.
5) WildWINGS- a fun activity on Friday that offers skills development.
6) Kids Praise Programs- practicing and demonstrating ne found skills.

2.2 Characteristics of families, children, neighborhoods and schools

The schools and communities in North Charleston that are served by WINGS have high levels of social, economic and academic risk. Over 90% of students are black and over 90% are eligible for free or reduced-price lunch. The median family income in 2008 was $39,653, compared with $63,211 for the nation, placing the majority of North Charleston's residents below 200% of the poverty level (U.S. Census Bureau, 2008). Fifty-two percent of North Charleston births in 2008 were to single mothers. Given the incidence of crime relative to the population, North Charleston has been ranked the seventh-most dangerous city in the United States (Paras, 2007). Across the four schools served by WINGS, the majority of students (42% for reading, 52% for writing, 50% for math, 65% for science) do not meet statewide proficiency standards. The graduation rate for the high school attended by students in WINGS schools was 34.3% in 2007-08, compared to 73.2% for the nation as a whole (Cataldi, Laird, & Kewal Ramani, 2009; McGinley, Rose, & Donnelly, 2009).

2.3 The organization, staff, and their training

2.3.1 Organization

At the start of the study, WINGS operated in four schools in Charleston, but over the course of the study expanded to two additional cities (Atlanta and Charlotte). In order to support its multi-city operations, the WINGS organization (see Appendix C for organization chart) consists of staff at three major levels: central corporate; city; and school. Central corporate staff consists of the chief executive officer, executive staff that oversees major organizational functions and three executive directors that oversee operations in each of the three cities (and their surrounding areas). The executive directors, in turn, oversee the work of the regional directors, who are the direct link to each of the program schools.

The chief program officer and her staff are responsible for planning and facilitating the annual summer training for WINGS leaders in Charleston. She also works with staff from the corporate and regional offices to observe WINGS nests systematically and use the information to design coaching and training for WINGS leaders.

Regional managers are responsible for supervising the program directors at each of the schools in their region. They report directly to the executive directors, but are also accountable to the chief program officer. They plan and facilitate local trainings,

observe program directors and program assistants, and are primarily responsible for hiring WINGS leaders. They also work with program directors to identify ways to help WINGS leaders who are struggling and children who may be struggling within the program.

Program directors, program assistants, peace managers, and WINGS leaders comprise the staff working directly with children in the WINGS program. Program directors and program assistants work full-time, while peace managers and WINGS leaders work part-time.

In general, program directors have previous experience as WINGS leaders, although that varies across the regions. Each program director is responsible for program oversight at a school. They assist the regional director in hiring and training WINGS leaders, and they are responsible for coaching their WINGS leaders and providing training over the course of the school year as needed.

Program directors also facilitate Community Unity, which is typically the first session of every day. Community Unity is a time for children to sit down with their "nest," or the group of children they usually spend their afternoon with, eat snack, say the "WINGS Creed," which summarizes the social-emotional goals for the WINGS program, and go over/reinforce the specific social-emotional objective for the week. Critically, Community Unity marks the transition from the regular school day to the WINGS after-school day, and it requires that the program director be an energetic, well-organized leader with very strong social skills who is well liked and respected by children.

Program assistants support the program directors. They do not provide training and oversight to the WINGS leaders, but they manage logistics at the school, assist with program planning, and step in to watch children if a WINGS leader needs to step away from the children in his/her nest.

Peace managers support WINGS leaders in managing the behaviors of WINGS children who may come into the program tired from the school day, upset about events at home or with friends, or otherwise unsettled. If a child acts out, and the WINGS leader is unable to redirect him/her from disrupting the group, then the WINGS leader may ask the WINGS peace manager to talk with the child or provide a quiet place where the child can calm down before rejoining his/her group. Peace managers may also step in to watch a nest of children if the WINGS leader needs to be absent from the room.

WINGS leaders work directly with a small group of children, called a "nest." Nests are intended to be 12 or fewer children. They are typically organized by grade level and, in some cases, gender. Thus, a school may have a first-grade girls' nest or a first-grade boys' nest. If it has more than 24 first graders, then it may have a third nest that is mixed gender. At times, the WINGS staff will mix children from two consecutive grades

and have a kindergarten/first grade nest. By creating these configurations, program directors work to ensure that no more than 12 children are in a nest, a target that is typically achieved.

WINGS for Kids requires that WINGS leaders–those young adults who work directly with children–be college students because it wants to emphasize and model the importance of education for children. Also, because the program places a high priority on the development of relationships between WINGS leaders and children, WINGS leaders are expected to work five days a week, from approximately 2 to 6 p.m.

2.3.2 WINGS Leaders and their Training

The study conducted an initial survey of WINGS nest leaders early in the study to provide more information on their characteristics, experience, attitudes, and job climate (see Appendix D for the survey instrument, response rate and preliminary analysis). WINGS leaders tend to be recruited from the college student population, and work for WINGS on a part-time basis. The typical WINGS leader works five days a week for about four hours a day. WINGS leaders may or may not come into the program with previous experience with children, and, if they have experience, it varies a lot from leader to leader. A minority report that they have worked with children in camps; others report that they have younger relatives they have cared for.

Turnover can be high among WINGS leaders, and program supervisors report that WINGS leaders who leave during the first year often do so because the demands of WINGS and the demands of their school and home lives conflict. For instance, during the 2015-2016 school year, about 2/3 of the WINGS leaders were in their first year. Formal training of WINGS leaders occurs in the summer of each year, but training is ongoing throughout the year and includes both formal training and on-the-job training.

Formal Training at the Beginning of the Year- Prior to Summer 2015, WINGS leaders received one week of summer training which tended to focus heavily on program activities and logistics. However, the program staff added an extra week of training (five half days) that allowed more focus on handling behavioral issues.

The first week, which consisted of all-day sessions held jointly for all the Charleston area schools, focused on child development and the need to support, engage, and teach children (GET SET), learning through play, and managing children's behavior in constructive ways (such as teachable moments, diverting children's attention, and disregarding some negative behavior). WINGS leaders were introduced to a number of interactive activities that reinforce social-emotional learning objectives.

The second week of training, held separately at each school for five half days by the program director, focused on program logistics: the calendar, daily activities, transitions between activities, dismissal at the end of the day, interactions with teachers, and the

use and care of school spaces. This week provided opportunities for program directors and their new WINGS leaders to get to know one another.

__Formal trainings During the School Year-__ In addition to the summer trainings, WINGS leaders receive formal training during the school year from the regional operations directors and/or their program directors. These trainings are often on specific topics that pose challenges for WINGS leaders. For example, one training held in 2015-2016 in Charleston was "Attention Getters," which was intended to teach WINGS leaders how to get children's attention without giving them direct commands and, hopefully, making the process more fun and engaging for the children. The content of the trainings are determined by the program directors and program assistants in conversation with regional directors. As part of their ongoing professional development, the program directors and assistants are encouraged to conduct the trainings with their staff, although the regional director may design the training.

__On-the-job Training-__ WINGS leaders hired during the school year typically undergo on-the-job-training. Ideally, they will be paired with a more experienced WINGS leader and given the opportunity to shadow and work with that person's guidance. However, when a WINGS leader leaves the program with little or no notice, new WINGS leaders occasionally find themselves on the job with little training.

__Coaching-__ All WINGS leaders receive frequent coaching from program directors and, to a lesser extent, program assistants. Both program directors and assistants spend much of their afternoons helping staff with their nests and managing logistical issues that arise, which provides many opportunities for them to provide feedback to WINGS leaders. According to WINGS leaders surveyed in Spring 2016, about three-quarters received feedback from their supervisors at least once a week, and most feedback was provided verbally as the WINGS leaders and their managers went about their daily activities.

In addition to the coaching they receive from their program directors, corporate staff from the central office observes each WINGS leader at least once during the year, and the information from the observation is used to shape future trainings. For example, observations led corporate staff to conclude that children's movement was being overly restricted (e.g., children were expected to sit in seats, instead of being allowed to sit on the floor), which led to changes in the WINGS leader training.

2.4 WINGS attendance and retention (across all grades)

2.4.1 Enrollment and Attendance at WINGS

This section examines enrollment and attendance patterns at WINGS for all the children in the Charleston study schools (grades K-5). Later in the report, we will discuss attendance patterns for the study sample. However, since that sample has somewhat

different enrollment and attendance patterns from the overall population of students who go to WINGS, it is helpful to describe attendance in the program overall.

WINGS tracks children's attendance and children are expected to attend almost every day. If they miss too many days in a row, program staff will reach out to them to find out why. If they continue to miss WINGS, then staff may ask them to leave the program (severe and consistent misbehavior is another reason for asking a child to leave the program). This attendance policy is motivated both by wanting the program to impact as many children as possible, and wanting each child to receive significant dosage to obtain the effects of the program.

However, enrollment and attendance data are mainly affected by the high rate of family migration for low-income urban families and the stress and turbulence in their lives. These topics are described in more detail in chapters 4 & 5 where their effect on the study sample is discussed.

2.4.2 Enrollment by year at study schools

Table 2.1 shows WINGS enrollment for the study years by school. WINGS serves about 500 children in Charleston a year and three schools (Chicora, Memminger and NCES) provide the largest share of the students. JSE provided a much smaller share and the WINGS program was terminated in the final year due to a shift in the demographic characteristics caused by JSE becoming a Montessori School.

Table 2.1 WINGS Enrollment Across Years by Study School, All Grades

School	2012-2013	2013-2014	2014-2015
Chicora	166	125	159
Memminger	139	127	132
James Simons	75	115	N/A[1]
NCES	176	144	159
TOTAL	556	511	440

[1] After the 2013-2014 school year, JSE transitioned to a Montessori school and the characteristics of the student population changed from low- to middle-income. WINGS terminated its program at JSE in the final study year.

Table 2.2 presents average daily attendance rates for children enrolled in the program by grade for all schools for two school years. Attendance rates are similar across years. Attendance rates average about 80 percent, but attendance for kindergarten and first grade children tends to be lower (about 75%). This lower rate may reflect that young children may encounter more developmental challenges from the very long, extended school day- typically from 8-5:30.

Table 2.2 WINGS Attendance Rates by Grade
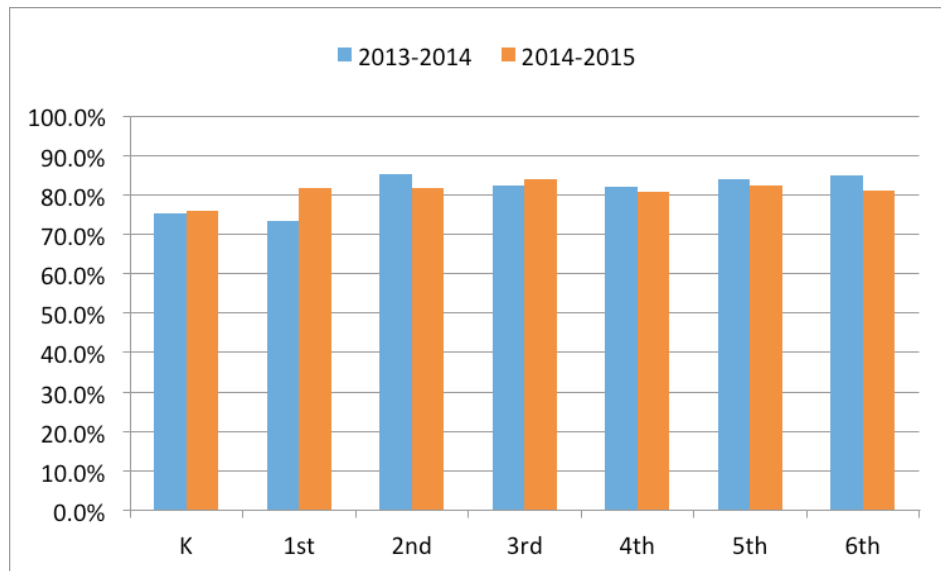
Chapter 3   Literature Review and Research Design

3.1 Literature Review

3.1.1 The importance of socio-emotional skills

Poverty is a powerful condition that can alter the trajectory of children's cognitive, physical, and socio-emotional development (Duncan & Magnuson, 2005; Evans & Kim, 2013). In recent years, the increase in child poverty has led to greater attention to resilience–the ability to succeed in the face of risk or challenge (Oades-Sese, Kaliski & Weiss, 2010)–and has sparked a commitment to understanding the intrinsic, protective characteristics of children who grow up in poverty (Whittaker, Harden, See, Meisch, & T'Pring, 2011). Over the years, numerous studies have pointed to social-emotional competence, a multi-faceted construct that consists of skills that enables one to handle social tasks in appropriate ways, as an important mechanism of psychosocial resilience.

However, many children struggle with mastering social-emotional competence. Children from low-income families, in particular, face many challenges and risks related to their social-emotional development that can have negative consequences later on in life (Duncan & Magnuson, 2005). Unfortunately, gaps in social-emotional development between low-income children and their more affluent peers are observed before entering formal schooling, and these gaps persist and increase during the elementary school years and beyond (Alexander, Entwisle, & Kabbani, 2001; Brooks-Gunn, Duncan, & Aber, 1997; Hamre & Pianta, 2001). Without early intervention in social-emotional and behavioral skills, young children are at greater risk for future academic problems, dropping out of school, peer rejection, and antisocial behaviors (Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011; Durlak & Weissberg, 2011; Greenberg et al., 2003). On the other hand, strengthening young children's social-emotional competence may serve as an important protective factor for school and life success, especially if they are exposed to multiple life stressors (Jones, Greenberg, & Crowley, 2015; Webster-Stratton, Reid, & Hammond, 2004). As such, researchers have begun investigating promising approaches and intervention programs, ranging from in-school curriculums to teacher and parent training programs, that target the promotion of social-emotional competence in children (e.g., Jones & Bouffard, 2012; Morris et al., 2013; Morris et al., 2014; Webster-Stratton,Reid, & Hammond, 2004). There is less work, though, on the impact of social-emotional competence interventions in after-school settings, despite the fact that these competencies can be taught in various ways across many different settings and contexts (CASEL, 2016).

3.1.2 Children's social-emotional competence

Social-emotional learning (SEL) broadly refers to the process by which cognitive, affective, and behavioral skills are acquired which help children effectively establish and maintain positive, healthy relationships, successfully carry out various social tasks, and meet daily challenges (CASEL, 2016; Denham et al., 2012; Nickerson & Fishman, 2009). In young children, being socially and emotionally competent means that they are able to inhibit impulsive behavioral responses, take into account others' perspectives, make good decisions, express healthy emotions, recognize problems and provide feasible solutions, and adjust and integrate emotions, behaviors, and actions, in order to work well socially with others, act responsibly and respectfully, and display developmentally appropriate prosocial behaviors (Denham et al., 2012; Durlak et al., 2011; Weissberg, Caplan, & Sivo, 1989; Zins, Elias, Greenberg, & Weissberg, 2000). Although there are many ways to consider SEL, five core components of SEL have been identified (CASEL, 2016) and are specifically highlighted within this report: self-awareness, self-management, responsible decision-making, social awareness, and relationships skills.

Self-awareness captures the ability to accurately recognize one's feelings, attributes, and values, and understand how those feelings influence behavior (Denham & Brown, 2010). For young children, learning new words to label how they feel and describe what led to those feelings, and developing a sense of self, including knowing what they like and dislike and identifying strengths and weaknesses, are important for developing self-awareness (CASEL, 2016). *Self-management* describes the ability to successfully regulate one's emotions, thoughts, and behaviors, and appropriately express them in multiple contexts, as well as the ability to manage stress, control impulses, and set goals and persist in meeting those goals (CASEL, 2016). Although some children may be able to describe how they are feeling, most children transitioning to formal school are still learning how to express and react to their feelings and match them to the expectations of the different situations and contexts they encounter. *Responsible decision making* entails learning how to make constructive and respectful choices about personal behavior and social interactions, analyzing and solving problems, being ethically responsible, and considering the well-being of oneself and others (CASEL, 2016; Denham et al., 2010). With the help of adults, children are learning how to make choices based on personal opinions, social norms, and rules, and contemplating the consequences of their actions. *Social awareness* is the ability to understand what behaviors are socially and ethically acceptable in different situations and contexts, as well as the ability to take another's perspective, and show empathy toward others, including those from diverse backgrounds (CASEL, 2016; Denham et al., 2010). Through interacting with peers and adults, children from a young age are learning how to interpret others' emotions and understanding that how they feel may not necessarily be how others are feeling. *Relationship skills* refer to the ability to establish and maintain

healthy relationships across a diverse range of individuals, as well as to utilize skills, such as cooperation, listening, negotiating, seeking and offering help when needed, and turn taking, to build and sustain these relationships (CASEL, 2016; Denham & Brown, 2010). Young children are beginning to learn what it means to "be a good friend", ask for and offer help, communicate effectively, cooperate, negotiate conflicts, and share.

Each of these components, though distinct, is highly interrelated. Both individually and together, they predict a range of positive outcomes (CASEL, 2016; Denham et al., 2012). For instance, children's social-emotional competence has been linked to positive relationship skills and behaviors, both in the classroom and at home, and increases in children's long-term academic skills. Importantly, the role that contextual factors play in children's social-emotional competence and related skills also needs to be considered, especially when examining children from low-income, high-risk backgrounds.

3.1.3 Building blocks of social-emotional competence

Many interventions and programs targeting the promotion of social-emotional competence aim to promote the building blocks that set a strong foundation for social-emotional development in young children (e.g., Morris et al., 2014). Executive function (EF), in particular, has received much attention given its critical role in the development of social-emotional competence (e.g., Bierman, Nix, Greenberg, Blair, & Domitrovich, 2008; Blair & Raver, 2015; Riggs, Jahromi, Razza, Dillworth-Bart, & Mueller, 2006).

Executive function. Executive function (EF) is a multi-faceted construct that can broadly be defined as the processes of cognitive flexibility, working memory, and inhibitory control that are necessary for purposeful, goal-directed behavior. Studies show that there are persistent and growing poverty-related gaps, not only in achievement (Reardon, 2011), but also in the regulation of attention, emotion, stress response, and executive function (Cicchetti, 2002; Evans, 2003). Evidence from neuroscientific studies suggest that focusing on EF can enhance children's learning and development and can establish positive academic trajectories, particularly for children from low-income families (Blair & Raver, 2015; Evans & Schamberg, 2009; Raver, Blair, Willoughby, & Family Life Project Key Investigators, 2013).

Specific to children's social-emotional competence, EF has been directly implicated in the concurrent and longitudinal development of social-emotional skills (e.g., Riggs, Blair, & Greenberg, 2003). This is not surprising given the many overlaps between the subskills that underlie both EF and social-emotional development. Studies shows that difficulties in EF lead to difficulties in multiple components of social-emotional functioning, including impulsivity, delay of gratification, emotion regulation, problems with attention, behavioral issues, and problem solving (e.g., Cole, Usher, & Cargo, 1993; Hughes, 2002; Jahromi & Stifter, 2008; Kim et al., 2016; Pennington, 2002; Seguin, Boulerice, Harden, Tremblay, & Pihl, 1999). Moreover, the EF components related to

planning, inhibiting response, and controlling one's attention may be particularly useful for resisting temptation, regulating frustration and stress (Mischel et al., 1989), and behaving according to social norms (Peake, Hebl, & Mischel, 2002).

3.1.4 Social-emotional competence and children's academic and social skills

Learning is a social process, where students learn in collaboration with their peers, teachers, and families. Therefore, it is not surprising that social-emotional competence has been linked not only to other social skills but also to academic outcomes (Denham, 2006; Denham et al., 2012; Durlak et al., 2011; Zins, Bloodworth, Weissberg, & Walberg, 2007). In fact, in early childhood, SEL is especially critical for social and academic success. Social-emotional competence in preschool predicts academic success in first grade, even after controlling for IQ and economic risk (Denham et al., 2012). Studies demonstrate that children who are not able to demonstrate age appropriate social-emotional competence participate less in class, are less accepted by their peers and teachers, and get fewer instructions and less positive feedback from elementary school teachers (Denham et al., 2012; Raver, 2002; Shores & Wehby, 1999). In addition, children who exhibit aggressive and antisocial behaviors tend to perform poorly on academic tasks, be held back or drop out of school, and struggle to maintain healthy relationships (Raver, 2002).

Although SEL instruction focuses mainly on social-emotional competencies, a main objective of these programs is to improve long-term academic performance. Being able to manage emotions, solve problems, and work cooperatively with others may contribute to success in school (Durlak et al., 2011; Elias, Zins, Graczyk, & Weissberg, 2003). Greater self-awareness and confidence around learning abilities is linked with increased student motivation, while improved goal-setting, stress management, and organizational skills are related to higher achievement (Greenberg et al., 2003). Improved social relationships, especially with teachers, are associated with greater school engagement and higher achievement, as well (CASEL, 2016). There is variation in the extent to which SEL programs explicitly address academic achievement, but those that promote the integration of SEL with academic curricula and teaching are increasing in number (Zins et al., 2004). There is some evidence to suggest that the presence of an academic component (such as tutoring or homework help) in an SEL program is a strong predictor of the program having a significant academic effect (Durlak & Weissberg, 2007).

3.1.5 Evidence of intervention impacts on social and academic outcomes

Unfortunately, it is more difficult to conduct research on the effects of socio-emotional and other non-cognitive skills on academic outcomes, than it is to study the direct improvement of academic outcomes by way of cognitive skills. The lack of routine data collection on socio-emotional skills, the small number of programs developing these

skills, and the challenges of working within natural settings has resulted in a research literature that is methodologically limited (Durlak, Weissberg, & Pachan, 2010; Lauer et al., 2006). Research that is mostly non-experimental suggests that SEL programming promotes positive youth development across a wide developmental span; in school-based, after-school, and community settings; with students who do and do not have presenting problems; in urban, suburban, and rural areas; among racially and ethnically diverse student bodies; and as implemented by professionals as well as school staff (Payton et al., 2008; SEL Research Group/CASEL, 2010). Studies have suggested students benefit from SEL across a wide range of outcomes, including having higher quality relationships with peers and adults, having fewer problem behaviors at school, using drugs and alcohol less, engaging in risky sexual behavior less, and behaving violently less (Greenberg et al., 2003). Students also have better attitudes about themselves, others, and school, and earn higher grades and test scores (Durlak, Weissberg, et al., 2011; Payton et al., 2008).

In a meta-analysis of 317 studies of SEL programs, Payton et al. (2008) suggested that SEL programming was associated with students' gaining an average of 11 to 17 percentile points on achievement tests. Among the 180 studies of programs considered "universal" (not targeted), the authors found a mean effect on academic performance of .28. Effects on other outcomes such as attitudes toward self and others, positive social behavior, conduct problems, and emotional distress were similarly in the .20 range (Payton et al., 2008). Similarly, in another meta-analysis of 213 studies involving more than 270,000 students, Durlak and colleagues (2011) found that, overall, SEL programs both in and out of school were significantly effective (grand study-level mean = 0.30). Specifically, students who participated in evidence-based SEL programs demonstrated enhanced SEL skills, better attitudes about themselves, others, and school, and increased prosocial behaviors, compared to students who did not participate in these programs. Students participating in the SEL programs also had fewer conduct and internalizing problems, lower levels of emotional distress, increased ability to manage stress and depression, as well as significant gains of 11 percentile points in academic achievement compared to students in the control group. Follow-up data (at least six months later) showed sustained effects in all areas listed above, with effect sizes ranging from 0.11 to 0.32.

3.1.6 Family contextual factors related to social-emotional competence
The family context is particularly critical in young children's lives given the rapid development that occurs during this early period and the important role that parents play in children's development. Families from low-socioeconomic backgrounds encounter a myriad of disadvantages that can range from living in harsh social environments to less employment opportunities and quality support services for parents and family members (Duncan, 1991; Zigler, 1994). These poverty-related factors can have negative

consequences for children's development and overall success (McLoyd, 1998; Sektnan et al., 2010; Shonkoff & Phillips, 2000). For children from impoverished backgrounds, positive interactions with peers and adults support healthy development and learning (Birch & Ladd, 1998; Hamre & Pianta, 2007; Ladd & Burgess, 2001). However, there are many different aspects that can contribute to the interactions that occur within and outside of the learning environment. Parent stress and mobility are two important factors that can have specific influences on children's social-emotional competence.

Parent stress. According to cumulative risk models, exposure to multiple risk factors, such as poverty, parent stress, mobility, and low parent education, can create increased stress for children and families (Evans & Kim, 2007), which can have long-term detrimental effects (Blair & Raver, 2012). Stress in the home environment is an important mechanism that can partially explain the link between family adversity, such as economic hardship and instability, and negative outcomes (Blair & Raver, 2012).

Mobility. Residential mobility is a relatively common occurrence for low-income families living in the United States (U.S. Census, 2011). Residential mobility is typically thought of as the number of moves or instability in housing that families experience and can be an important indicator of the home environment quality (Evans & Wachs, 2010). Compared to their more affluent peers, children from low-income families move nearly two times more often (Cohen & Wardrip, 2011). This is especially critical given that residential mobility is strongly related to young children's development and well-being (Blair & Raver, 2012). Research shows residential mobility is negatively related to a wide range of outcomes, including physical health (Busacker & Kasehagen, 2012), social-emotional development (Brown, Ackerman, & Moore, 2013), self-regulation (Schmitt, Finders, & McClelland, 2015), relationships (Haynie, South, & Bose, 2006), negative behaviors (Ziol-Guest & McKenna, 2014), and academic achievement (Herbers et al., 2012; Schmitt et al., 2015; Schmitt & Lipscomb, 2016; Voight, Shinn, & Nation, 2012).

Potential mechanisms underlying the negative relation between residential mobility and children's outcomes may include family stress and chaos. Although mobility can signify increased opportunities for families (i.e., upward mobility), more often than not, frequent moves are likely to be associated with stress, disruptions in routines and social networks, housing hazards, crowding, and general instability (Coulton, Theodos, & Turner, 2012; Leventhal & Newman, 2010). Another potential mechanism may be that moving frequently, especially prior to beginning formal school, may mean that children are not prepared to enter kindergarten, which can have long-term implications for their academic and behavioral trajectories (Obradovic, 2010; Ziol-Guest & McKenna, 2014). Interestingly, research suggests that there may be a threshold effect in the number of moves. In one study, mobility was a not a significant factor in predicting school readiness outcomes for families who moved one to two times; however, moving three or

more times was significantly related to poor attention and problem behaviors, suggesting that multiple and frequent disruptions due to housing is problematic for children's development (Ziol-Guest & McKenna, 2014). However, recent evidence suggests that children's executive functions, specifically inhibitory control, may help to buffer the detrimental effects of residential mobility on children's outcomes and may be an important skill to target in early interventions and programs for young children experiencing multiple environmental risks (Schmitt et al., 2015).

3.1.7 Existing interventions and programs targeting social-emotional competence

Our current understanding of child development has been strongly influenced by a set of experimental and quasi-experimental early interventions implemented in the 1960s to '80s that included long-term follow up of the participants into adulthood (Campbell & Ramey, 1995; Campbell, Ramey, Pungello, Sparling, & Miller-Johnson, 2002; Finn & Achilles, 1999; Gray & Klaus, 1970; Karoly et al., 1998; Reynolds et al., 2007; Ramey et al., 2000; Schweinhart, 2004). These interventions continue to be the focus of re-analysis, partly to address issues with small sample sizes, and to assess whether results are robust across studies (Anderson, 2008; Barnett, 1995; Krueger & Whitmore, 2001). Results have confirmed the importance of intervening early and have identified a broad range of long-term academic, social, and economic effects that can be achieved with early intervention.

The focus of these early interventions varied somewhat, with some primarily addressing academic outcomes, and others targeting a wider range of outcomes. However, the results across interventions, including a recent re-analysis of three experiments (Abecedarian, Perry Preschool, and Tennessee Early Training projects) utilizing new estimation techniques appropriate for very small samples, showed that early academic gains (e.g. in achievement test scores) that were initially quite large generally decayed over time. This resulted in most, if not all, of the gains disappearing by middle school (Anderson, 2008; Barnett, 1995; Krueger & Whitmore, 2001). However, despite this decay, there remained substantial long-term academic, social, and economic benefits. These benefits included higher high school graduation and college entrance rates; higher income from higher wages and lower unemployment; and less drug and alcohol addiction, teen pregnancy, welfare utilization, and criminal justice involvement. Interestingly, the analysis by Anderson (2008) that combined data from the three experiments found that these long-term benefits were confined almost entirely to girls. This further suggests that larger sample sizes are important for gender sub-group analysis.

Cost-benefit analyses of several early interventions point to significant investment returns to the individuals involved and to society as well, with many studies showing that the major financial benefits derive from changes in youth behavior during adolescence,

rather than higher long-term adult incomes (Karoly et al., 1998; Karoly et al., 2006; Lynch, 2004; Masse & Barnett, 2002; Reynolds, Temple, Robertson, & Mann, 2002). These results suggest that students' improved short-term academic outcomes may not be the driving force behind these financial gains. Rather, it has been hypothesized that improved "non-cognitive" skills, such as socio-emotional skills, may be responsible for a significant part of the long-term returns. Research in economics over the past few years has also suggested that these non-cognitive skills may predict higher educational attainment and long-term income as strongly, if not more strongly, than direct achievement gains (Heckman, 2008; Heckman & Rubinstein, 2001; Bowles, Gintis, & Osborne, 2001; Bowles, Gintis, & Osborne, 2002). Additionally, Grissmer and Eiseman (2008) provide evidence that a substantial part of later achievement gaps are present prior to kindergarten entrance, and that gaps in non-cognitive skills between minority and disadvantaged students and their more advantaged peers may account for a significant part of early achievement gaps.

Social and emotional learning (SEL) programs gained traction in the mid 1990s as a coordinated and comprehensive counterpoint to fragmented school-based interventions that were widely implemented but seldom shown to be effective (Greenberg et al., 2003). The goal of SEL programs is to enhance youth competence and reduce risk of poor outcomes by building skills such as recognizing and managing emotions, understanding others' perspectives, goal-setting, responsible decision-making, and effective interpersonal communication (Greenberg et al., 2003; Zins, Bloodworth, Weissberg, & Walberg, 2004). *These skills are thought to enable students to develop more positive social relationships and ways of relating to themselves, others, and tasks, which in turn likely assist their efforts to learn in the school context.*

3.1.8 Characteristics of successful SEL programs

Few SEL programs have been evaluated using experimental methods with long-term follow-up. However, enough empirical evidence exists to suggest that certain types of programs are more effective in improving student outcomes than others. Greenberg et al. (2003) described the most effective programs as those that use structured manuals and curricula to create consistency in program delivery; address a range of SEL constructs; are long-term, with multi-year programs being best; and provide a developmentally appropriate progression of opportunities for skill-building, spanning ideally from pre-kindergarten to adolescence. Zins et al. (2004) noted effective programs tend to be theory- and research-based, with the most rigorous undertaking continuous self-improvement through outcome evaluation.

The Collaborative for Academic, Social, and Emotional Learning (CASEL), a group at the forefront of SEL research and theory development, asserts that a combination of social competency instruction and positive learning environments (e.g., a safe and

supportive school climate, active partnership between family and school) contributes to children's short- and long-term success (http://casel.org/sel/academics.php; Zins et al., 2004). A number of SEL programs based on this theory, including WINGS, utilize a model of instruction built around a framework of five key person-centered SEL competencies - self-awareness, social awareness, responsible decision-making, self-management, and relationship management (Payton et al., 2008; Zins et al., 2004). Improvements in these skill areas, in conjunction with positive environments, are hypothesized to lead to less risky behavior, greater attachment to school, better academic performance, and more success in life (http://casel.org/sel/academics.php).

Results from a recent series of meta-analyses of SEL program effects further suggest that theory-based programs that go on to employ evidence-based skill-training approaches in social competency instruction are the most effective (Durlak & Weissberg, 2007; Durlak, Weissberg, et al., 2010; Payton et al., 2008). *More specifically, SEL programs that provide training that is sequenced, active, focused, and explicit (given the acronym SAFE) have greater effects on student outcomes across a number of domains.* Notably, in these meta- analyses, when effect sizes were calculated separately for programs that met SAFE criteria and programs that did not, in many domains where there had previously been a significant overall effect, the effects of non-SAFE programs fell to non-significance while the effects of SAFE programs remained (Durlak & Weissberg, 2007; Durlak, Weissberg, et al., 2011; Payton et al., 2008). For example, in their review of 180 studies of universal SEL programs, Payton et al. (2008) found an overall effect size on students' positive social behavior of .24. However, when differentiated, the mean effect size of SAFE programs was .28, while the mean effect size of non-SAFE programs was .02.

Durlak and Weissberg (2007) and a follow-on meta-analysis by Durlak, Weissberg et al., (2011) found SAFE SEL programs had significant positive effects on a range of student outcomes, including child self-perceptions (self-confidence and self-esteem), school bonding, positive social behaviors, problem behaviors, drug use, achievement test scores, school grades, and attendance. Average effect sizes across these outcomes ranged from .14 (school attendance) to .37 (child self-perceptions) (Durlak & Weissberg, 2007; Durlak, Weissberg et al., 2010). *Taken together, these results suggest that SEL programs that meet SAFE criteria have particular promise as an intervention promoting positive youth development. As described below, the WINGS program uses a skill-training approach that qualifies as SAFE.*

3.1.9 After-school programs

After-school programs (ASPs) provide children with supervision and care between the end of the school day and the end of most parents' workday. They have become increasingly common and are in high demand as maternal employment and the number of single-parent families in the US has grown over the years (Durlak, Mahoney, Bohnert,

& Parente, 2010; Gottfredson, Gerstenblith, Soule, Womer & Lu, 2004). Up to 8.4 million school-aged youth (15% of the school-aged population) are currently estimated to attend ASPs (Durlak, Mahoney et al., 2010; http://afterschoolalliance.org), while another 15.1 million (26%) are estimated to be unsupervised (http://afterschoolalliance.org). Research has shown that lack of supervision during the after-school hours is related to increased association with delinquent peers, and to higher rates of juvenile crime and delinquent behaviors (Gottfredson et al., 2004; Riggs & Greenberg, 2004). Recognizing the need for after-school programming, the US Department of Education offered nearly one billion dollars in ASP funding in 2009, with state and local governments and private sources contributing additionally (Durlak, Mahoney et al., 2010). Rigorous evaluations of these programs are needed to better utilize the significant investment being made on federal, state, and local levels.

Evidence of impacts on youth development. Studies have shown participation in after-school programming is associated with positive social, emotional, and academic outcomes for children (Riggs & Greenberg, 2004). ASP participation is related to increased school attendance, school engagement, grades, test scores, graduation rates, and parental worker productivity. It is also related to decreased drug and alcohol use, truancy, school dropout, sexual activity, juvenile crime, and obesity (Durlak, Mahoney et al., 2010; http://afterschoolalliance.org). However, while promising, these results remain preliminary for a number of reasons. Findings have been inconsistent across studies and programs, and research methods often lacked rigor (i.e., not randomized, no comparison group, etc) (Kane, 2004; Lauer et al., 2006; Riggs & Greenberg, 2004). Additionally, youth often participate in multiple out-of-school time activities, so it is difficult to attribute a child's progress to any one program (Durlak, Mahoney et al., 2010).

In their meta-analysis of 27 studies of after-school programs, McComb and Scott-Little (2003) found the evidence regarding ASP effects on students' academic achievement to be inconclusive. There was some indication, however, that lower-achieving students may gain more from ASP attendance than higher-achieving students, and that students who have higher ASP attendance also benefit more (McComb & Scott-Little, 2003). In a separate review, Kane (2004) found that ASP participants may earn higher grades and do homework more consistently than non-participants, even if they show no increase in achievement test scores. Additional study of after-school programs, employing rigorous methods, is necessary to better understand the full extent and range of their effects on child outcomes.

Characteristics of successful after-school programs. In assessing the effectiveness of after-school programs, it is important to consider their heterogeneous nature. There are generally two types of ASPs: childcare programs and youth development programs (Riggs & Greenberg, 2004). Childcare programs provide care

but advance no other specific goals, whereas youth development programs contain purposeful programming that explicitly attempts to promote positive development in any number of areas. These areas can include academics, social and emotional skills, and physical fitness (Riggs & Greenberg, 2004). Studies of ASP effectiveness are generally only concerned with youth development programs.

The specific elements of youth development programming and program climate are two factors that have been associated with participant outcomes (Riggs & Greenberg, 2004). One study by Gottfredson et al. (2004) showed that among all types of youth development ASPs, programs that place a high emphasis on social skills and character development have the greatest effect in reducing students' delinquent behaviors. Similarly, Fashola (1998) found that opportunities for one-on-one tutoring, and a strong link between ASP activities and the school-day curriculum, were the most promising strategies for improving academic achievement. However, findings from Lauer et al. (2006) suggest that after-school programs need not focus solely on academics in order to produce academic effects. In a meta-analysis of 42 ASP studies that collected reading outcomes and 33 studies that collected math outcomes, the authors found small but significant effects of after-school programming on both math and reading. *They also found that programs that had a combined academic and social focus - like the WINGS after-school program - had a larger average effect size on math than programs that had a primarily academic focus (.19 compared to .07) (Lauer et al., 2006).*

Program climate characteristics that are known to be related to overall ASP effectiveness include program quality, implementation quality, and levels of child participation, among others (Durlak, Mahoney et al., 2010; Hirsch, Mekinda & Stawicki, 2010; Riggs & Greenberg, 2004). Elements of program quality associated with positive youth outcomes include high-quality staff, positive staff-child relationships, and rich and varied programming. Indicators of high implementation quality, such as well-trained staff, efficient program procedures, and low child-to-staff ratios are also related to improved youth outcomes (Durlak, Mahoney et al., 2010; Riggs & Greenberg, 2004). Greater child program participation (enrollment, attendance, and engagement) is also related to improved youth outcomes (Hirsch et al., 2010).

Added value of SEL after-school programs. A valuable hybrid of an after-school program (ASP) and a social and emotional learning (SEL) program is the after-school social and emotional learning program. These are SEL programs that take place in the after-school context. SEL ASPs have the potential to be particularly effective interventions, especially for low-resource children, since they provide supervision during the after-school hours, in addition to social and emotional competency instruction. Also, unlike school-based SEL programs, they are able to operate freely without the constraints of the school-day structure and curriculum.

Research has shown SEL ASPs that utilize skill-training approaches that are SAFE (sequenced, active, focused, explicit) have better outcomes in improving children's social, emotional, and academic functioning. In their meta-analyses of studies evaluating after-school SEL programs, Durlak and Weissberg (2007) and Durlak, Weissberg et al. (2010) found that SAFE SEL ASPs have effects on child outcomes that are equal to or greater than the effects of other types of commonly-seen interventions for youth, such as drug prevention or mentoring programs. These authors also found the mean effect size of SAFE SEL ASPs was .20 - .31 for students' achievement test scores, .22 - .24 for school grades, and .16 - .37 for a range of other outcomes including self-perceptions, school bonding, positive social behaviors, problem behaviors, and drug use (Durlak & Weissberg, 2007; Durlak, Weissberg et al., 2010). *Taken together, this evidence suggests that the WINGS program, as an SEL ASP that meets SAFE criteria, is likely to be an effective program.*

Figure 3.1 depicts the WINGS program's theory of change. The stated objectives of the program are to enhance children's person-centered competencies (self awareness, social awareness, responsible decision-making, self management, and relationship management), and the theory of change follows that (a) children assigned to WINGS will develop greater person-centered competencies than children who are not assigned to WINGS. The theory of change also proposes that the positive impacts of assignment to WINGS will transfer to the classroom and home setting. Specifically, we hypothesize that compared to children not assigned to WINGS, children assigned to WINGS will develop and display more positive and less negative behavior and relationships with their teachers and classmates in school and at home. The (b) increased positive and decreased negative relationships and behaviors are hypothesized to be (c) mediated through the improvements in children's enhanced person-centered competencies. Finally, assignment to WINGS is proposed to have (d) longer term positive impacts on children's academic school outcomes and socio-emotional behavior. We also collected a broader set of measures for exploratory analysis that focused on measures of early emotional and cognitive skills that can contribute to identifying possible causative mechanisms that underlie the impacts as well as interpret the pattern of results across outcome measures.

Figure 3 .1 Theory of Change
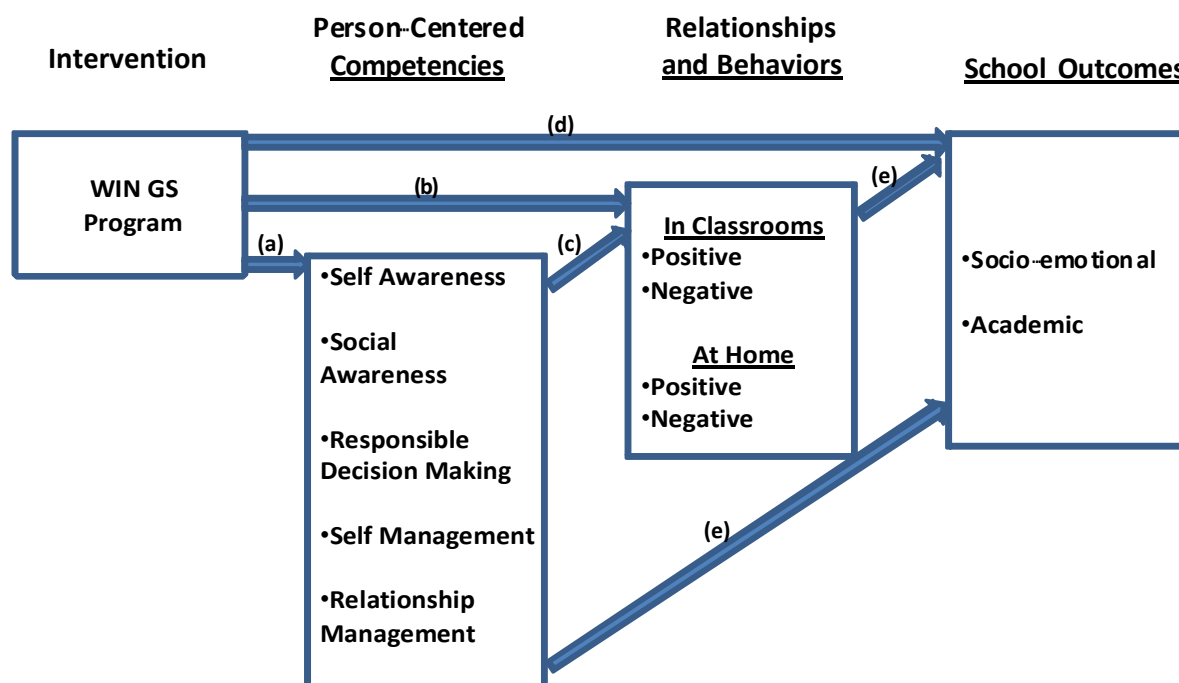
3.2 Research Questions

The following research questions can be addressed by the study.

- *Direct Effects and Mediated Effects of Random Assignment to WINGS (Intent-to-Treat)*

  - *Question 1. Does assignment to WINGS have a positive impact on children's person-centered competencies?*

  - *Question 2. Does assignment to WINGS have a positive impact on children's relationships and behaviors in the classroom and at home?*

  - *Question 3. Does assignment to WINGS have a positive impact on children's short-term school academic outcomes?*

  - *Question 4. What are the impacts of WINGS on children's person-centered competencies, and school and home outcomes after one year and two years of WINGS participation?*

- *Moderated Effects of WINGS Related to Child Characteristics and Site*

  - *Question 5. Does the impact of WINGS on children's person-centered competencies, and relationships and behaviors at school and home after one year and two years vary for children with different characteristics?*

  - *Question 6. Does the impact of WINGS on children's person-centered competencies, and relationships and behaviors at school and home vary across the different sites within which random assignment was conducted?*

  - *Question 7. Does the impact of WINGS on children's person-centered competencies, and relationships and behaviors at school and home change from cohort 1 to cohort 2?*

  - *Question 8. Does the impact of WINGS on children's person-centered competencies, relationships and behaviors, and school outcomes vary by the level of initial skills?*

In addition, there are several exploratory questions linked to the measures included in our Building Block Skills. These questions include whether WINGS has an impact on each skill, and whether these skills mediate the main impacts.

3.3 Research Design

Proposed Design The original design outlined in the SIF proposal was a plan for a four cohort study- two cohorts to be funded by IES (N = 260) and two cohorts to be funded by SIF (N = 180). This original research design proposed a *four-site, four-cohort, block randomized control trial*, in which entering kindergartners would be randomly assigned to 32 implementation blocks. Specifically, child-level random assignment to WINGS or control will be determined within four schools and within each of four entering cohorts of eligible kindergartners who sign up to enroll in the program (i.e., schools and cohorts within schools will serve as randomization blocks). Because the program serves 12 girls and 12 boys who enter kindergarten each year and conducts social-emotional learning activities separately within each gender "nest", gender will also serve as a randomization block as boys and girls were each randomly assigned separately. Thus there would be 4 cohorts, 4 schools and 2 genders for 32 implementation blocks. This design addressed most common threats to internal validity, and results would provide strong evidence about the causal impacts of random assignment to WINGS.

Research Design for Current Evaluation This study evaluates two cohorts who receive up to two years of WINGS participation. Cohort 1 had four schools and cohort 2 had only three schools due to the discontinuation of the program at James Simons. James Simons transitioned to a Montessori Magnet school that changed its demographic characteristics, and the WINGS program was discontinued. Specifically, child-level random assignment to WINGS or control will be determined within four schools in cohort 1 and three schools in cohort 2. Because the program serves 12 girls and 12 boys who enter kindergarten each year and conducts social-emotional learning activities separately within each gender "nest", gender will also serve as a randomization block to ensure equal numbers of girls and boys are enrolled in the program. Thus, there will be 14 blocks for the current two cohort study evaluating two years of WINGS.

Recruitment. Within each school, 24 kindergartners (12 boys and 12 girls) were assigned to participate in WINGS from among a list of entering kindergartners whose parents/guardians enrolled their child. Through word-of-mouth and informal referral processes, the WINGS program documents that each school has full enrollment in the program and a wait-list for enrollment of, on average, 10 kindergartners (ranging from 6-12). Active recruitment strategies were used to increase the total number of children who are eligible to participate in this RCT. These included providing opportunities for enrollment at spring kindergarten registration and fall kindergarten orientation, and sending WINGS program materials and enrollment forms home with parents on the first day of kindergarten. In addition, the fee for WINGS enrollment is paid by private donations, so there will be no cost to families. In sum, we projected a total of 36

entering kindergartners who will enroll for the program each year, and 24 will be assigned to participate in WINGS and 12 assigned to the control group.

<u>Sibling Preference.</u> Preference for enrollment in WINGS was given to kindergartners who have siblings already enrolled in the program, and the WINGS program documents that this accounts for approximately 25% of children who are awarded a slot (6 of the 24 slots per school). Because of the non-random assignment to the program of these six children, they were not included in the randomized study testing the program's impacts. Randomization was conducted for the 18 remaining slots per school among the 30 remaining eligible kindergartners.

<u>Random Assignment and Informed Consent.</u> Participation in the study was voluntary. We sought informed consent to participate in this study from the parent/guardian of each of the 30 kindergartners who are registered to enroll in WINGS and who were not awarded a slot based on sibling preference. We randomly assigned these 30 eligible kindergartners (whether parental consent is received or not) to the 18 remaining WINGS slots within each school. To randomly assign children, we will collect lists of kindergartners eligible for random assignment and generate a randomly ordered list for each school. From these lists we will select children for WINGS, beginning at the top of the list and moving down, selecting as many children as there are slots available. Randomization will occur separately by gender and will conclude after 12 children have been assigned to WINGS within each gender block "nest." We will then provide the lists of accepted children to the schools, so they may offer enrollment to these children. The children who remain unselected will be placed in the control group. Following these procedures, we estimate that for each cohort within each school, 18 kindergartners will be randomly assigned to WINGS and 12 kindergartners to the control group.

Chapter 4 Major Data Collections, Timelines and Response Rates

The study tracked each cohort from kindergarten entrance through 2nd grade entrance in order to allow children two years of possible WINGS participation (see Table 1.1). Pre-test data was collected in the summer/fall of kindergarten entrance and post-test data for one year of potential WINGS participation was collected in the spring of kindergarten and the summer/fall following kindergarten. Post-test data for two potential years of WINGS participation was collected in the spring of 1st grade and in the summer/fall following first grade.

Table 4.1. Percent Data Collected (Response Rates) by Time Point, Measure Type, and Cohort.

Time Point	Type of Measure	Cohort 1	Cohort 2
Summer/Fall of Kindergarten	Direct Measures	99%	98%
	Parent Reports	92%	97%
	Teacher Reports	93%	96%
Spring of Kindergarten	Teacher Reports	96%	97%
Summer/Fall of First Grade	Direct Measures	95%	84%
	Parent Reports	89%	75%
	Teacher Reports	86%	53%
Spring of First Grade	Teacher Reports	73%	60%
Summer/Fall of Second Grade	Direct Measures	77%	63%
	Parent Reports	73%	63%

Table 4.1 shows the response rates for each type of assessment at each time point by cohort. Due in large part of the mobility of study families, the prevalence of missing data increased throughout the course of this study, with more missing data occurring for each cohort in later time points than in early time points. For instance, Cohort 1 response rates at baseline (Summer/Fall of Kindergarten) were above 90% for all types of measures, but were less than 80% at the end of 1st Grade and beginning of 2nd Grade.

Studies of children that occur outside of the regular attendance during schools days encounter challenges in tracking and collecting data from parents and testing children. Low-income, urban families are a particular issue due to their frequent relocation of households and changing schools for their children. Table 4.2 shows the percentage of children that were enrolled in a non-study school by the summer of 2015, approximately three years after the start of the study for Cohort 1 and two years after the start of the study for Cohort 2. This data suggests an annual migration of 20% of children relocating to non-study schools during the study. Study children that originally attended four study schools are currently dispersed across at least 52 different schools, only ten of which are outside of South Carolina. Part of the cause of the relocation can be changes in jobs or income that demand a move. The higher relocation rates at Memminger and NCES suggest that some relocation may have been to better housing and/or jobs since parents had higher education levels than at Chicora.

Table 4 . 2 Percent of Children Relocated by Cohort and School.

School attended at start of study	Percent relocated to non-study schools as of Summer, 2015	
	Cohort 1	Cohort 2
Chicora	39%	29%
Memminger	68%	43%
NCES	68%	34%
JSE	43%	N/A
Total	56%	35%

Migration to non-study schools causes problems in both tracking and maintaining contact with families, but also requires a new relationship be established with the new schools so that teacher surveys can be administered, and, in some cases, testing of children can be done during or after school. Overall these moves can increase the cost of data collection significantly due to the extra time and travel required, but they also increase the attrition rates as parents and teachers have less knowledge and incentive to continue participation.

Missing data can be problematic in RCTs in two ways. The first concern is if the characteristics of participants with data differ from those without data. The second concern is that missing data reduces the statistical power of the sample to detect effects. A particular concern is when there are different levels of missing data in test and control groups which suggests that the cause of missing data may be different between

test and control groups, causing potential bias in estimated effects. In the next section, we assess differential attrition between test and control groups. More broadly, we address missing data by estimating ITT and TOT effects and compare results from a sample with complete data to the sample including all cases with missing data with imputed values (see Chapter 7-RESULTS). The results show little sensitivity to the effects of missing data. In the next chapter, we provide a more detailed analysis of the causes of attrition and potential effects on biasing impact estimates.

Chapter 5 Sample Characteristics, Attrition and Non-Compliance

This chapter provides descriptive statistics regarding differences between test and control groups for the initial baseline sample, and then focuses on the differential attrition and non-compliance for test and control groups at later time points for the sample receiving one year of treatment and the sample receiving two years of treatment.

5.1 Sample Randomization

Table 5.1 shows that children were randomized within 14 randomization blocks: two cohorts, at four schools, by gender group (Female/Male), with one school (James Simons) having only one cohort. Across Cohorts 1 and 2, 101 children (41.7%) were assigned to the control group and 141 (58.3%) were assigned to the treatment group. Table 5.2 also provides the sample sizes for each randomization block, as well as the probabilities for being assigned to treatment or control within each block. Overall, about 60% of participants were assigned to treatment with 40% to the control group. We assigned more to treatment in order to compensate for expected non-compliance.

Table 5 . 1 Randomization Blocks – Frequencies and Probabilities.

			Random Assignment - Frequencies (n)		Random Assignment - Probabilities	
Cohort	School	Gender	Control	Treatment	Control	Treatment
1	Chicora	Female	11	16	40.7%	59.3%
		Male	5	9	35.7%	64.3%
	Memminger	Female	6	9	40.0%	60.0%
		Male	9	13	40.9%	59.1%
	NCES	Female	9	13	40.9%	59.1%
		Male	8	11	42.1%	57.9%
	James Simons	Female	6	7	46.2%	53.8%
		Male	4	4	50.0%	50.0%
2	Chicora	Female	8	12	40.0%	60.0%

		Male	6	8	42.9%	57.1%
	Memminger	Female	7	10	41.2%	58.8%
		Male	6	7	46.2%	53.8%
	NCES	Female	7	11	38.9%	61.1%
		Male	9	11	45.0%	55.0%
TOTAL			101	141	41.7%	58.3%

5.2 Sample Demographics

Table 5.2 provides descriptive statistics for sample demographic characteristics and life circumstances as measured at the beginning of the study. These factors are important because such background characteristics can affect whether and how much children will benefit from a particular intervention. As expected, children in this sample would be characterized as living in "high-risk" circumstances, with a majority (96%) qualifying for free or reduced-price lunch and a majority having racial minority status (91% Black, 5% Latino/a). Further, the overall education level of mothers was low, with more than a quarter (29% of respondent caregivers) having less than a high school degree, and unemployment was high (35% of respondent caregivers). Many caregivers also report major changes prior to participating in the study: half (49%) moved during the two years prior to the study child entering kindergarten, more than one in six (17%) moved more than once in that time; half (46%) reported a major change in employment or work-related responsibilities, hours, or conditions; and more than a third (36%) reported someone in the home starting or stopping either work or education. These characteristics suggest that these low-income families frequently encountered a major change in life circumstances that could affect their level of stress, the behavior and development of their children, and participation and attendance at WINGS.

Table 5 . 2 Sample Demographic Characteristics.

Baseline Characteristic or Experience	Mean (SD)	%
Child Gender (% Male)		46.8
Adult Race (% Black)		92.2
Receiving Free/Reduced Lunch		95.7
Receiving Other Forms of Public Assistance		80.8
Parent Employed (or Student)		64.4
Attended Preschool		88.5
Mother's Education		
Less than HS		27.1
HS/Equivalent		37.2
More than HS		35.6
Age (years, on first day of school)	5.5 (0.3)	
Mother's Age (years, at time of first questionnaire)	29.0 (4.8)	
Number of Children in Home	2.9 (1.4)	
Number of Adults in Home	1.5 (0.7)	
Number of Moves in 2 yrs prior to K	0.8 (1.0)	
Holmes-Rahe Life Stress Inventory weighted score	249.5 (167.2)	
Perceived Financial Strain (possible range: 1-5)	2.0 (0.9)	
Perceived General Stress (possible range: 1-5)	2.6 (0.6)	
Parent-Child Relationship Stress (possible range: 1-5)	1.5 (0.4)	

5.3 Differential Attrition and its Significance

Differences in response or attrition rates between test and control groups can be problematic and pose a potential source of bias in effect estimates depending on whether the characteristics of those students with data are similar to those without data. Table 5.3 provides response rates by major data collections (parent, teacher, and child testing) by treatment and control group for the one year of participation sample and the two year of participation sample. The table also shows the tests for the statistical significance of the differences. Only 1 of 12 comparisons between treatment and control

groups shows a significant difference. Only two of twelve comparisons do not meet conservative WWC attrition and differential attrition standards, and all comparisons meet the liberal WWC standards.

Table 5 . 3 . Attrition by Time Point, Measure Type, and Treatment Group.

Time Point	Type of Measure	Percent Collected		Difference
One Year of Participation Sample				
		Treatment (n = 209)	Control (n = 145)	% Diff (*p*-value)
Summer/Fall of Kindergarten	Direct Measures	96%	94%	2% (*p* = .62)
	Parent Reports	92%	82%	10% (*p* < .01)
	Teacher Reports	87%	90%	-3% (*p* = .44)
Spring of Kindergarten	Teacher Reports	88%	94%	-6% (*p* = .28)
Summer/Fall of First Grade entry	Direct Measures	81%	85%	-4% (*p* = .35)
	Parent Reports	79%	74%	5% (*p* = .32)
Two Years of Participation sample				
		Treatment (n=141)	Control (n=101)	% Diff (p-value)
Fall of First Grade	Teacher Reports	70%	77%	-7% (*p* = .31)
Spring of First Grade	Classroom Observations	65%	67%	-2% (*p* = .75)
	Teacher Reports	67%	67%	0% (*p* = 1)
Summer/Fall of Second Grade	Direct Measures	70%	71%	-1% (*p* = .97)
	Parent Reports	68%	69%	-1% (*p* = .95)

The significant differences was present only for parent reports at baseline, with a 92% response rate for the treatment group and an 82% response rate for the control group (10% difference, *p* < .01). This may have been due to parents/caregivers of children who were randomized into the control condition feeling less of an obligation or desire to respond to requests for baseline data. Generally speaking, though, because treatment groups did not differ in their extent of missing data at later time points, we would not

expect that missing data should be a significant source of bias in estimating treatment impacts.

There were almost no significant differences between treatment and control on demographic characteristics reported in the quantitative questionnaire at baseline (see Table 5.4). This suggests that random assignment was balanced on demographic characteristics. The only differences between treatment and control groups were in number of children in the home and occurrence of stressful life events in the year prior to the start of the study. Treatment children had, on average, 2.9 children living in their homes, compared with 2.5 children, on average, living in control children's homes ($p < .05$). Our metric for occurrence of stressful life events (as measured on the Holmes-Rahe Life Stress Inventory) is a weighted score based on weights developed by the measure's authors in order to adjust for severity of each event (e.g., a death of the caregiver's spouse is weighted 100 points, whereas a major change in eating habits is weighted only 15 points). On average, the weighted score for stressful life events for treatment caregivers was 249.5, compared to 193.6 for control caregivers ($p < .05$).

Table 5 . 4 . Sample Demographic Characteristics by Treatment and Control.
(Characteristics ordered from most different to least different between treatment and control.)

Baseline Characteristic or Experience	Treatment		Control		Comparison	
	Mean (SD)	%	Mean (SD)	%	Difference	p val.
# (%) Respondents	*n* = 209	92.3	*n* = 145	82.1	10.2	0.003
Number of Children in Home	2.9 (1.4)		2.5 (1.3)		0.4	0.02
Holmes-Rahe Life Stress Inventory weighted score	249.5 (167.2)		193.6 (133.7)		54.9	0.04
Attended Preschool		88.5		92.4	-3.9	0.33
Perceived Financial Strain (possible range: 1-5)	2.0 (0.9)		2.0 (0.9)		0	0.40
Adult Race (% Black)		92.2		89.0	3.2	0.41
Number of Moves in 2 yrs prior to K	0.8 (1.0)		0.7 (0.8)		0.2	0.42
Mother's Ed						0.60
Less than HS		27.1		32.5	-5.4	
HS/Equiv		37.2		34.2	3.0	
More than HS		35.6		33.3	-2.3	
Age (years)	5.5 (0.3)		5.5 (0.3)		0	0.70
Receiving Free/Reduced Lunch		95.7		96.6	0.9	0.77
Child Gender (% Male)		46.8		48.2	1.4	0.83
Perceived General Stress (possible range: 1-5)	2.6 (0.6)		2.6 (0.6)		0	0.85
Receiving Other Forms of Public Assistance		80.8		79.8	1.0	0.88
Number (#) of Adults in Home	1.5 (0.7)		1.5 (0.6)		0	0.92
Parent-Child Relationship Stress (possible range: 1-5)	1.5 (0.4)		1.5 (0.4)		0	0.95
Mother's Age (years, at start of study)	29.0		29.1		0.1	0.96

Because these differences existed between treatment and control groups at baseline, we used number of children in the home and parental stress levels as covariates in our analyses. This is necessary in order to ensure that the two randomly assigned groups are "balanced," or comparable, a necessity in order to interpret post-intervention differences as causal effects of the WINGS program. We also did sensitivity analyses to the inclusion and exclusion of covariates in our analysis, and results (see Chapter 7-RESULTS) indicate little sensitivity to the inclusion or exclusion of covariates.

5.4 Sample Non-Compliance

In contrast to the term "study attrition," which refers to individual children or families not being available to provide data for the *study*, the term non-compliance refers to whether the participants complied with their treatment and control assignment. Participants who won the lottery are non-compliers if they do not meet the standards for completing either one or two years of WINGS (No-Shows). Participants who lost the lottery are non-compliers if they actually attend the WINGS program and meet the standards for one and two years of attendance. These participants are termed crossovers. Non-compliance can introduce bias if the non-compliers' characteristics are not similar to compliers for both treatment and control groups. Non-compliance also is problematic because it reduces the statistical power of the sample.

According to the WINGS logic model, attending WINGS for two years is thought to be necessary before seeing positive impacts. In conjunction with WINGS personnel, we also established minimum attendance criteria for each year based on actual attendance data. We set the criteria of at least 100 days of attendance in kindergarten and first grade to qualify as having "received treatment." Based on these criteria, Tables 5.6 and 5.7 provide the consort data for the level of compliance and non-compliance for Cohort 1 and Cohort 2 respectively.

Table 5.6 shows that only 30 of 82 (37%) participants assigned to treatment in Cohort 1 met the attendance criteria in both kindergarten and 1st grade, while the compliance rate for treatment in Cohort 2 was 42%. The compliance rates for one year of attendance were much higher with 68% (Cohort 1) and 61% (Cohort 2) of treatment children complying. The compliance rates for control children were very high with 86% (Cohort 1) and 91% (Cohort 2) of children not receiving treatment.

Table 5 . 6 Cohort 1 Two- Year Consort Data

Randomized Children:	140							
Treatment Condition:	Treatment (n=82)				Control (n=58)			
Attended at least 100 days in Kindergarten?	Yes		No		Yes		No	
	56		26		1		57	
% of condition group	68%		32%		2%		98%	
Attended at least 100 days in First Grade?	Yes	No	Yes	No	Yes	No	Yes	No
	30	26	8	18	0	1	7	50
% of condition group	37%	32%	10%	22%	0%	2%	12%	86%

Table 5 . 7 Cohort 2 Two- Year Consort Data

Randomized Children:	102							
Treatment Condition:	Treatment (n=59)				Control (n=43)			
Attended at least 100 days in Kindergarten?	Yes		No		Yes		No	
	36		23		1		42	
% of condition group	61%		39%		2%		98%	
Attended at least 100 days in First Grade?	Yes	No	Yes	No	Yes	No	Yes	No
	25	11	5	18	1	0	3	39
% of condition group	42%	19%	8%	31%	2%	0%	7%	91%

The study maintained records for each child that drew from parent conversations and WINGS personnel and recorded the reasons for treatment children withdrawing from the program. Table 5.8 shows these reasons for non-compliance. About 60% of non-compliance for treatment children was due to a relocation and attendance at another school without the WINGS program. The WINGS program was available only in the four study schools, so almost all relocation was to district schools without WINGS. The second most important reason for non-compliance (23%) was removal by the parent without relocation. Removal by the parent could occur for a wide number of reasons and might reflect the increased stress on the child and parent from the longer day at school, and the parent preference to have the child return home after school. Parents respond very positively when asked about the WINGS program, and there is little evidence that dissatisfaction with WINGS is a significant cause of removal.

Table 5 . 8 Reasons Reported for No Longer Attending WINGS

Reason Reported	Cohort 1	Cohort 2	%
Relocated	27	19	60%
Removed by family (other than relocation)	10	8	23%
None/unknown reason	3	1	5%
Removed by WINGS	3	2	6%
Attending other program	1	0	1%
Health issues	3	0	4%
Total	47	30	100%

The threshold of 100 days as delineating a full year of WINGS attendance was one suggested by WINGS personnel in the spring of 2015, but is not the only plausible definition for "receiving treatment." A stricter definition would require a greater number of days of attendance (e.g., 120 days) in both kindergarten and first grade. More lenient definitions include simply attending a total of 200 days or 240 across kindergarten and first grade. For example, according to the "total of 200 days" definition, a child could be counted as having received treatment if she attended 150 days in kindergarten and 50 days in first grade.

Table 5.9 provides the total compliance rates according to these four different definitions of treatment receipt by cohort. Note that compliance rates are lowest overall for the strictest definition of treatment (i.e., at least 120 days attended in both kindergarten and first grade). They are not necessarily the lowest for the most lenient definition of treatment (i.e., at least 200 days across kindergarten and first grade) because this definition allows for more children in the control group to be counted as "receiving treatment" than, for example, defining treatment as at least 240 days across both grades. These compliance rates come into play when understanding the difference and relationship between ITT impact estimates and treatment-on-treated (TOT) impact estimates. TOT estimates are substantially higher than ITT estimates because the ITT estimates include all children assigned to the program, while TOT estimates include only children who actually attend WINGS for the required dosage.

Table 5 . 9 Compliance Rates Using Various Definitions of Treatment Receipt

Definition of Treatment Receipt	Cohort 1	Cohort 2	Cohorts 1, 2
At least 120 days in *both* kindergarten and first grade	31%	41%	35%
At least 100 days in *both* kindergarten and first grade	37%	40%	38%
At least 240 days across kindergarten and first grade	32%	49%	39%
At least 200 days across kindergarten and first grade	34%	42%	37%

Chapter 6   Outcome Measures and Analysis Plan

First, we describe our outcome measures that flow from our theory of action and research questions described in earlier sections. Then we provide the analytical approach to estimation of the impacts of WINGS including handling of missing data. We also estimate the interactions by gender, cohort, school, and level of initial skill.

6.1 Outcome Measures

Assessment tools used in this study included direct child assessments, and measures from teacher and parent reports on classroom and home behavior and relationships. Appendix E contains the detailed characteristics of the outcome measures and their characteristics. Table E.1 provides a list of 38 outcome measures grouped by three categories (Child Assessments, Teacher Reported, and Parent Reported) and by our theory of action categories (Building Blocks, Teacher Positive Behaviors, Teacher Negative Behaviors, Parent Positive Behaviors, Parent Negative Behaviors, and Academic). The description and psychometric properties of these 38 measures are provided in Table E.2. All outcome measures obtained from item-level responses on the parent or teacher surveys described below are contained in Appendix F.

Person-Centered Competencies. Direct assessments were completed in areas that align closely with the constructs of self-awareness, social awareness, responsible decision-making, self-management, and relationship management identified in the theory of change. Parents and teachers reported on the five SEL skills (self-awareness, social awareness, responsible decision-making, self-management, and relationship management) via the *Devereux Student Strengths Assessment* (DESSA; Lebuffe, Shapiro, & Naglieri, 2008).

Teacher-report measures of children's relationships and classroom behaviors included the *Student-Teacher Relationship Scale* (STRS; Pianta, 2001), which measures the quality of the teacher's relationship with individual children, and the *Social Skills Improvement System* (SSIS; Gresham & Elliott, 2008), which is a measure of an individual child's relationships and social behaviors in the classroom. Children's relationships and behaviors at home were assessed during parent/guardian interviews using parent versions of the *Social Skills Improvement System* (SSIS; Gresham & Elliott, 2008) and the *Child-Parent Relationship Scale* (CPRS; Pianta, 1992).

School Outcomes. Direct assessments of academic outcomes were completed using the *Woodcock-Johnson-III Tests of Achievement* (WJ-III; Woodcock; McGrew, & Mather, 2001), which evaluates reading skills (Sound Awareness and Letter-Word Identification subtests), mathematics skills (Applied Problems and Quantitative Concepts subtests), and general knowledge (Academic Knowledge).

Building Block Skills included measures of executive function: *Head-Toes-Knees-Shoulders Task* (HTKS, Ponitz, McClelland, et al., 2008), *Emotion Matching Task* (EMT; Morgan, Izard, & King, 2009), *Assessment of Children's Knowledge Task* (ACES; Mavroveli et al., 2009), *Theory of Mind* (NEPSY II; Korkman, Kirk & Kemp, 2007a, 2007b), *Delay of Gratification Task*, (Mischel, Shoda, & Rodriguez, 1989), and Visual-Motor Integration (VMI; Beery VMI; Beery, 2010; Beery, Buktenica & Beery, 1997).

6.2 Estimation Methodology

We estimated the ITT and TOT impacts of WINGS on all 38 measures for both one year and two years of participation using multiple imputation that incorporates the entire sample. The ITT analysis provides estimates of the impact on all children assigned to WINGS, regardless of whether they attended, while the TOT analyses provides estimates of the impact for children who actually attended. The TOT effects show much larger effects than ITT effects because only approximately four out of ten treatment children completed two years of WINGS.

A strength of experimental design is simplicity; the impact of a program can be estimated by simply comparing the mean outcomes of those assigned to the treatment group to the mean outcomes of the control group. However, experiments that have non-compliance, missing data, differences in measure reliability, multiple outcome measures, and can encounter improbable random draws require a more robust analysis that involves doing extensive sensitivity analysis of the results to a variety of assumptions.

In this analysis, we provide estimates for two distinct samples: the sample that includes only cases with complete data (list-wise deletion) and the full sample with imputations for all missing data. We also provide results for three different specifications. The first specification contains the treatment variable with no covariates, while the second specification adds the pre-test as a covariate. The third specification adds a list of additional covariates to control for variations in random assignment and variations in the time between pre- and post-test and whether the test occurred in the summer or during the school year. Our summary tables in Chapter 7 contain six estimates for each of the outcome measures for both ITT and TOT estimates. The six estimates include 3 specifications (no covariates, pre-test only and full set of covariates) for list-wise (cases with complete data) and multiple imputation (full sample with all missing data imputed). Full regression results are available upon request. Generally the results and their interpretation are insensitive to the inclusion of additional covariates and to imputation of missing data. We also discuss the pros and cons of the issue posed by multiple comparisons in Chapter 8.

In this report, we present and discuss the results with full imputation of missing data including all covariates.

Intention-to-Treat (ITT) Analysis. The ITT estimates are calculated from a multiple regression model:

$$Y_i = \beta_1 T_i + \sum_{q=1}^{6} \delta_q X_{qi} + \varepsilon_i$$

where Y_i represents an achievement outcome Y for student i, T_i is a dichotomous variable indicating the lottery outcome for student i (0 = lottery loser-control; 1 = lottery winner-treatment), and the coefficient β_1 is the WINGS treatment effect of interest. X_{qi} is a vector of covariates, including the pre-test, dummies for each of the schools, a set of demographic variables (gender, age, mother's education, number of children in home, age of mother at birth), and variables for controlling differences in time between pre-test and post-test and whether testing occurred in the summer or school year. The latter variables were adjusted for the differences in the number of days between pre- and post-tests and also for whether tests were administered during the summer or during the school year.

Child gender, cohort, and school were each included in the covariates list since these were variables on which random assignment was blocked. Including child age improves precision by removing variance in child outcomes due to age/maturation. Inclusion of child pre-test scores allows for interpretation of impact estimates as being on *development* or *improvement* in child outcomes. Mother's education and age at child's birth were included as they were related to many child outcomes – as tends to be the case in the literature – and therefore also improve precision of estimates. Number of children in the home was not balanced between the assigned treatment and control groups; therefore, it was necessary to include this variable as a covariate to achieve balance between randomly assigned groups. Finally, mother's age at child's birth was included as it was related to number of children in the home (with older mothers more likely to have more children in their home) and is also related to child outcomes; therefore, its inclusion as a covariate made interpretation of the covariates' coefficients more easily interpretable.

For parent reports and direct assessments, we also included test-date adjustments for both pre-test and post-test measures. These adjustments (measured in days since the start of school, with a negative entry indicating that an assessment took place during the summer) are necessary in order to improve precision, since children's assessment timing is sometimes related to their outcomes. These adjustments were not possible with teacher reports, as all teacher reports were collected on roughly the same date; in contrast, the parent reports and direct assessments had wide windows for administration (from summer through fall).

For parent reports only, we also included a weighted score for instances of stressful life events, as reported at pretest (Holmes-Rahe Life Stress Inventory). This was because

random assignment was not balanced on this variable for Cohort 2 at baseline, and also because this measure was often closely related to parent-reported outcomes. Since this variable was typically not related to teacher reports or direct assessments, we did not include it in models for outcomes other than parent reports.

<u>Treatment-on-Treated (TOT) Analysis</u>. The ITT estimates present an internally valid estimate of the effect of being offered a position in WINGS. However, not all students who are assigned to the treatment group choose to enroll in WINGS and receive the required dosage of 100 days in kindergarten for the one-year impact analysis, or 100 days in both kindergarten and 1st grade for the two-year impact analysis. In addition, a few children who lost the lottery (crossovers) attended WINGS and received the required dosage. Therefore, the ITT estimates do not accurately measure the impact of actually *attending WINGS and receiving the required dosage*.

However, the choice of whether to attend WINGS is endogenous. In other words, it is possible that there is some unobserved reason, such as ability or motivation, why some students choose to attend and others do not. Models that fail to account for the endogeneity of this choice could suffer from selection bias if the unobserved reasons that influence their choice to attend are also correlated with their outcomes. Fortunately, whether or not a student is *admitted* to WINGS is exogenous–random and not influenced by any unobservable characteristics of the students. Our estimation strategy exploits the exogeneity of this assignment to attempt to estimate the causal relationship between WINGS attendance and achievement.

We estimate the effect of attending WINGS using an instrumental variables (IV) approach to account for "no shows," or those who choose not to attend after being offered admission, and "crossovers," or those who attend WINGS after not being offered admission (Bloom, 1984). In general, IV requires at least one instrument that is correlated with the endogenous choice (whether or not to attend) but is otherwise uncorrelated with the outcome (achievement). If students with an offer are more likely to attend, it satisfies the first requirement of an instrument. Because offers were random, we assume the lottery outcome is otherwise uncorrelated with outcomes.

The first stage equation, relating treatment assignment to treatment receipt, was as follows:

$$\hat{T}_i = \gamma_! + \gamma_! Z_i + \Sigma \gamma_! X_{i\cdot\cdot} \quad (2)$$

where $\hat{T}_i$ represents the expected treatment receipt status for child i (0 = did not receive WINGS treatment; 1 = received WINGS treatment), based on his/her treatment assignment and consideration of covariates, $\gamma_!$ is the effect of treatment assignment (Z_i; 0 = control; 1 = treatment) on treatment receipt (T_i), and X_{qi} is a vector of covariates with

their accompanying regression coefficients, γ_1, representing each covariate's "effect" on the child's likelihood to receive treatment.

The second stage equation, which models the association between predicted treatment receipt and child outcomes, is:

$$Y_i = \delta_0 + \delta_1 \hat{T}_i + \sum \delta_{qi} X_{qi} + e_i \text{, (3)}$$

where δ_1 represents the association between *predicted* treatment receipt ($\hat{T}_i$, as in equation (2)) and child outcome Y_i, and X_{qi} is our vector of covariates with their accompanying regression coefficients, δ_{qi}, representing each covariate's "effect" on the outcome. Because $\hat{T}_i$ is the predicted treatment receipt based on random assignment and covariates, the magnitude of δ_1 can be interpreted as the difference in child outcomes between children who do and do not receive WINGS treatment due to treatment and assignment and covariates. A reduced single-equation form of these analyses can be devised by substituting the right side of formula (2) in for $\hat{T}_i$ in formula (3). After some simplification, this reduced form can be written as:

$$Y_i = \beta_0 + \beta_1 Z_i + \sum \beta_{qi} X_{qi} + \varepsilon_i \text{, (4)}$$

where β_1 / δ_1 represents the estimate for our treatment effect.

6.3 Inclusion of Missing Data

The size of our randomized sample for two cohorts is 242 children, and we have complete data (listwise deletion) for all outcomes and covariates for154 children. Results using listwise deletion are viable only if attrition is random. Due in large part to the high residential mobility in this sample, study attrition (and missing data) increased over time throughout this study (see Chapter 5 for details on the source of missing data). One characteristic of the missing data is that it is seldom at the item level alone, but rather is due to an entire missing teacher survey, parent survey, or lack of all child assessments at least at one point in time. We have relatively complete data from all data sources at the pre-test, but due to migration, parent and teacher surveys and child assessments became increasingly more difficult to gather at later time points.

The extent of missing data by source of data (see chapter 4 and 5) meets, for the most part, conservative WWC standards, and all sources of attrition meet liberal WWC standards. However, the extent of missing data was large enough, and when combined with the levels of non-compliance and associated loss of statistical power, the listwise results alone would leave important stability issues. Thus, more complex methods needed to be incorporated into our analyses.

The inclusion of missing data into both the ITT and TOT estimates posed a challenge because the most frequently used procedure for incorporating missing data is full maximum likelihood estimates (FIML). FIML is considered a state-of-the-art missing data technique by many methodological experts (Enders, 2010) and is superior in its ability to reduce bias in statistical estimates in the presence of missing data that is not entirely random. FIML "borrows" information from all available data – for a particular observation and across all observations – in order to make a "best guess" at parameters defining relations among variables, even in cases where some observations are missing. By using all available information, as opposed to ignoring or removing cases with incomplete information, FIML preserves statistical power and produces less biased estimates than traditional alternatives (Enders, 2010).

FIML is often used by statistical software (or packages) which are based on a Structural Equations Model (SEM) framework. This framework has many advantages; for one, SEM allows for multiple outcomes to be included in the same model, thereby taking into account correlations among outcomes and reducing the number of statistical tests run (i.e., reducing the number of multiple comparisons, which reduces the chances of detecting spurious, significant effects). However, while FIML can be used for ITT estimates, we could not find any literature or software that allowed TOT estimates to incorporate FIML. TOT estimates are two-stage estimates, and FIML is currently only used for one-stage estimates. We utilized FIML in estimating ITT effects, and found that estimates were generally similar to estimates with listwise deletion (full data sample).

However, since TOT estimates are a critical component in interpreting the impacts of WINGS, we utilized a second procedure for handling missing data - Multiple Imputation (MI). MI can be utilized to produce both ITT and TOT estimates using the same assumptions and framework. MI is a modern and powerful missing data handling approach, and, like FIML, is also considered state-of-the-art (Enders, 2010). In contrast to FIML, which serves to provide the most probable parameter estimates that fit a given data set, the MI approach begins by filling in - i.e., imputing - probable values in place of missing data points. This process makes use of information from predictor variables as well as auxiliary variables (which do not appear in the model but are related to missingness in a particular variable and/or the variable itself), and also explicitly preserves uncertainty of imputed estimates. The imputation algorithm is repeated many times to create multiple imputed data sets, across which analyses can be run and then pooled to provide reliable parameter estimates. Thus, MI has the same benefits as FIML (reduced bias, preserved power), but also provides additional flexibility, as analyses do not need to take place within an SEM framework. Relevant to the current study, this flexibility allows for two-stage-least-squares estimation of TOT estimates.

MI is a general technique that can be carried out using any one of a variety of algorithms, procedures, and software tools. We used the MICE package (version 2.9) in

R (van Buuren & Groothuis-Oudshoorn, 2011), which is based on the multivariate imputation by chained equations (MICE) approach, alternatively called fully conditional specification. In a chained equations approach, the imputation model is defined variable by variable in order to allow for customization by analysts according to the unique features of each variable (e.g., the reasons why a particular variable has missing data). Then, each equation is run successively in an order defined by the analyst (in our case, predictors were imputed before outcomes) and the process is repeated over several iterations (in our case, 40 iterations) in order to improve precision.

Similar to the way we ran models in FIML, we imputed data separately by construct (or "bucket") represented in the WINGS logic model. Thus, a set of imputed data was generated for Teacher DESSA outcomes, along with all predictors in the Teacher DESSA models, for example. We chose to impute 40 data sets for each "bucket," and all subsequent analyses was performed on each of the 40 data sets before being pooled to produce final estimates. Auxiliary variables included child outcomes from "middle" time points, which were not included in the models (e.g., for "building blocks" outcomes in our two-year analyses, the kindergarten assessments were used as predictors, the second grade assessments as outcomes, and the first grade assessments as auxiliary variables).

Chapter 7   Results

The ITT results for one year of participation showed a overall pattern of null results with only a few reaching the p < .20 level of significance and none reaching the p < .05 level of significance. A supplementary spreadsheet with the results across measures is available upon request.

We report ITT and TOT impacts for two years of participation in WINGS for all 38 measures using the full sample and multiple imputation of all missing data. We report three levels of "statistical significance" at the p < .05, p < .10, and p < .20 level because lower levels of significance are important for providing guidance on improving the program, for interpreting patterns of results across measures, and possibly foreshadowing future effects with larger samples. We evaluate the magnitude of effect sizes by the conventional scale of small effect (~.25 SD), moderate (~.50 SD), and large (~.75 SD).

The two year ITT and TOT results with full covariates and multiple imputation for missing data are summarized below. The tables and figures showing these results are in Appendix G. Regression results are available upon request.s available upon request.

7.1 Teacher Assessment of Positive Behavior

Figure 7.1 (and Table 7.1) shows ITT effects for nine measures assessed for each child by teachers that measure a range of positive behaviors and skills. Eight of the nine measures show positive effects with effect sizes ranging from .08 to .20. Self management (effect size = .20, p < .20) and decision-making (effect size = .18, p < .20) show the strongest effects. Figure 7.2 (and Table 7.2) shows the TOT effects where the positive results for eight of nine measures range from .21 to .52. Self management (effect size = .52, p < .20) and decision-making (effect size = .42, p < .20) show the strongest effects in the moderate range.

7.2 Teacher Assessment of Negative Behavior

Figure 7.3 (and Table 7.3) shows ITT results for teacher assessment of six negative behaviors. All six measures show reduced negative behaviors with effect sizes ranging from -.12 to -.30. Four of the measures show significance of p < .10. These include hyperactivity (effect size = .31, p < .10), problem behaviors (effect size = .27, p < .10), bullying (effect size = .26, p < .10), and teacher conflict (effect size = .24, p < .10). Externalizing behavior (effect size = .23, p < .20) also shows reduced incidence. Figure 7.4 (and Table 7.4) shows the TOT estimates that are in the moderate to large range. These include hyperactivity (effect size = .77, p < .10), problem behaviors (effect size = .69, p < .10), bullying (effect size = .68, p< .10), teacher conflict (effect size = .63, p < .10) and externalizing behavior (effect size = .59, p < .20).

7.3 Parent Assessment of Positive Behavior

Figure 7.5 and 7.6 (and Tables 7.5 and 7.6) show ITT and TOT estimates for parent assessment of six positive behaviors. The ITT effects range from +.19 to -.18, but none reach the $p < .20$ level of significance. In contrast to teachers, parents report no significant change in positive behaviors for their children.

7.4 Parent Assessment of Negative Behavior

Figure 7.7 and 7.8 (and Tables 7.7 and 7.8) show ITT and TOT effects for parent reports on negative behavior. Parents report reduced or no change in negative behavior on all six ITT measures, but none reach the level of $p < .20$.

7.5 Building Block Skills from Direct Child Measures

Figure 7.9 and 7.10 (and Tables 7.9 and 7.10) show results for our category of building block skills. Two important skills show strong effects. HTKS (effect size = .26, $p < .10$) is a measure of executive function and DAS naming vocabulary (effect size = .26, $p < .05$) is a measure of verbal and comprehension skills. The corresponding TOT effects are in the moderate to large range: HTKS (effect size = .67, $p < .10$) and DAS naming vocabulary (effect size = .57, $p < .05$). None of the remaining six measures reach significance of $p < .20$.

7.6 Academic Skills from Direct Child Measures

Figure 7.11 and 7.12 (and Tables 7.11 and 7.12) show ITT and TOT results for our three academic measures. None of the academic measures show positive effects with significance $p < .20$. However, WJ Applied Problems shows a negative ITT effect (effect size = .21, $p < .20$).

7.7 Interaction Effects

The interactions by gender, school and initial skill level showed no consistent and significant patterns of effects across measures. However, a consistent cohort effect emerged across most measures that suggested WINGS impacts were larger in Cohort 2 than Cohort 1. Our mixed methods data also suggest that the quality of the WINGS program may have been adversely impacted for Cohort 1 both by geographical relocation of one of the schools and from WINGS program expansion to Atlanta and relocation of staff. In addition our data suggests that parent stress levels were much higher in cohort 1 than cohort 2 that may be linked to the school relocation and associated busing of child over longer distances.

Chapter 8   Discussion and Future Research Directions

8.1 Discussion of Results

The pattern and level of significance of the evaluation results for WINGS suggest that children attending WINGS in kindergarten and 1st grade have small to moderate impacts on children's positive classroom behavior and moderate to large impacts on reducing negative classroom behaviors. Many of these are in the $p < .20$ and $p < .10$ range, but none reach the $p < .05$ level. We also find that moderate to large gains on two important foundational cognitive skills of executive function and naming vocabulary at the $p < .10$ and $p < .05$ level, but the remaining six building block measures show no consistent effects reaching the $p < .20$ level.

These classroom behavioral effects might be linked to the effects for executive function and naming vocabulary. Execution function is predictive of less impulsive and more focused classroom behavior as well as improved social interactions. Improved social interactions might be linked to both less negative classroom behavior and improved vocabulary. The combination of improvement in two important foundational skills coupled with improved classroom behavior provides additional credibility to the results. Over 80% of measured interventions funded by IES find no significant effects, and those with effects are usually in the small to moderate range. Effects of this magnitude are rarely found in educational and social interventions. These effects are similar to or larger than effects measured in experimental evaluations of school-based socio-emotional programs. For instance, Blair and Raver, 2014 found effects of .3 to .8 for their low income sample on measures of EF and vocabulary. .

The absence of similar effects from parents and teachers suggest two hypotheses. The first hypothesis is that the child's classroom behavior has improved, but behavior at home has not improved. The second hypothesis is that parent ratings are biased compared to teacher ratings. Attendance at WINGS may make home behavior more challenging for children and parents. Our data shows that over 65% of control children who arrive home after school have parent/adult supervision with others attending after-school care programs. Home behavior may be different for children attending WINGS who arrive home much later after a long day at school and WINGS compared to children who arrive home after school and spend time with adults. We find that treatment mothers are more stressed than control mothers. This increased stress may result from the challenges of having a child attend WINGS and the associated challenges of late home arrival of a tired child. The classroom environment is also much more similar to and places similar demands on children as the WINGS environment, so transfer of new learned behavior during WINGS may be easier to the classroom than to the home environment.

Parent ratings of their own children can also be biased by the lack of a peer control group for comparisons that teachers have. Classroom behavior during the day for an entire school year provides an environment where a child's behavior can be compared to peers. Our data also suggest that parent ratings are strongly influenced by parent characteristics including the stress level of the parent. Parents of WINGS children, other things being equal, are more stressed, and stressed parents rate children's behavior lower.

The results also suggest that only attending WINGS in kindergarten does not produce these positive impacts found after two years. This pattern of impacts has three possible interpretations. One interpretation is that WINGS effects occur only if children attend for two years- a single year of attendance provides insufficient dosage for significant impacts. A second interpretation is that the WINGS program is not effective for kindergarten students, but the program is more effective for 1st grade students. A third interpretation is that the current two-year results partly reflect both longer term effects from kindergarten and short-term impacts from 1st grade. If effects are delayed, then effects will increase in the longer term.

Each of these interpretations would suggest that WINGS impacts would increase with more dosage, and/or with older children and/or if long-term impacts are measured. The WINGS program serves children through 5th grade, and the current evaluation does not include older children or children who receive more than two years of dosage or measure long-term results. Our evidence would suggest that effects may grow with more dosage, for older children, and may have delayed effects. Thus, the current evaluation that includes only younger children with limited dosage and measures short-term impacts may underestimate the full impact of the program.

WINGS is currently configured with only a few schools in a school district offering WINGS. For children who remain in schools that have the WINGS program, it is possible to attend up to six years since the program serves children through 5th grade. However, our data suggests two factors that make it difficult for children to receive higher levels of dosage. These factors are the high rate of migration of low-income families within school districts to schools not having a WINGS program, and the turbulence and stress present in these families from more frequent changes in jobs, income, health, and relationships that prevent regular attendance at WINGS. For instance, only four in ten children who were given the opportunity to start WINGS at kindergarten attained two years of treatment. Parents of WINGS children express strong approval and support for the program, but regular attendance requires remaining near schools offering WINGS and a level of family stability and commitment that some families find difficult to attain.

One direction for increasing the number of children who can have the opportunity to receive two or more years of WINGS from K-5 is to undertake a district-level

demonstration project making WINGS available in all schools in an urban school district. Doing so might substantially increase the opportunity for children to receive more than two years of dosage, and increase dosage for older students. Such a project would allow students who move within a school district to maintain WINGS attendance, and allow students who drop out of WINGS for a year or more to return and receive additional dosage. There are school districts throughout the nation who are experimenting with extended-day programs, but often utilize the extended time for more direct instruction. However, there is strong experimental evidence that more direct instruction in reading and math in extended-day programs does not lift achievement (Black, et al, 2009). However, the current evaluation results for WINGS appear strong enough to support a larger demonstration project in a school district that would allow experimental measurements of the long term impacts for older children and for children who receive more than two years of dosage.

8.2 Placing the Results in a Wider Context

Research on children's short- and long-term outcomes and the experimental evaluation of school-based programs and out-of-school programs to improve outcomes are undergoing rapid evolution and development. Historically, most of this research and evaluation was focused on the impact of schools and school-based interventions that have reading and/or math achievement as primary outcomes. School-based interventions using achievement as outcomes provide an advantage in evaluation because the measures are highly reliable and replicable, non-compliance and missing data are usually not problems, statistical power is high from large samples, and outcomes measures are few. In this evaluation environment, the standard ITT and TOT methods of evaluation using statistical significance levels of $p < .05$ are warranted. The best studies also incorporate multiple methods so that causative mechanisms can potentially be identified, and assessment can be done about whether some children benefit more than others and whether there are ways of improving the intervention through improved fidelity.

However, research is increasingly suggesting that children's long-term outcomes are predicted as much or more by developmental skills learned outside of direct school instruction than by skills learned in school direct instruction. These skills include self regulation, executive function, socio-emotional, visuo-spatial, and early comprehension. These early developmental skills have less reliable measures, require a larger set of outcome measures to capture their effects, and need more complex interventions for their improvement because they are learned largely outside schools. These interventions focused on activities outside of schools make large samples less accessible and have much higher levels of non-compliance and missing data that make adequate statistical power more difficult.

The measurement of these early developmental skills is a work in progress, and the quality and reliability of these measurements have not approached those of the most commonly used measures associated with achievement. Part of the problem is that any measure- achievement or otherwise- made with younger children have less quality and reproducibility. We should not expect at this stage of development to have the same kind of results that would be had using achievement measures because achievement measures are narrower, better developed – whereas these skills are displayed in a much wider set of behaviors, more difficult to measure, and less reproducible. These measures will improve over time and allow better measures- but at this stage of development, the criteria should not be the most stringent levels of statistical significance, i.e., $p < .05$. In the long run when measures and interventions are improved and samples have more power, imposing a statistical significance standard seems reasonable, but the purpose of an evaluation during the evolutionary period of improving measures and programs with weaker samples should incorporate a different set of objectives.

Evaluation methodology during the evolving period when measures and interventions are improving should incorporate the following elements:

- The standard RCT ITT and TOT analysis should be used that include methods of incorporating missing data with particular emphasis on the TOT effect sizes.
- Lower levels of statistical significance should be reported on measures
- Assessing ways of improving measures should be an important objective including incorporating a much wider range of exploratory outcome measures than is typical in RCTs.
- Assessing ways of improving the effectiveness of the intervention is also an important objective to undertake in addition to standard evaluation analysis.
- Interpretations should focus on the internal consistency and predictability from the theory of action of the broad patterns of results across measures rather than consideration of statistics across individual measures.
- No consensus exists about statistical adjustments for multiple comparisons when virtually all outcome measures are correlated. Any consideration of adjustments should recognize that while such adjustments decrease the chances of a false positive (an ineffective program will be labeled as effective), they increase the chances of a false negative (an effective program will be labeled ineffective). In the evolutionary research stages of developing improved measures and programs, searching for patterns linked to effective programs may be more important than guarding against misallocating resources to an ineffective program. Multiple adjustment comparisons do not generally shift the pattern of results that is the major focus of interpreting results during the early evolutionary stages. So they provide little useful additional information. As measures and programs improve and decisions are made involving significant resource allocations, tighter standards become more important.

Perhaps most importantly, the collection of mixed methods data seems critical for research involving the building of these early developmental skills in the out-of-school context. RCTs involving the building of early developing skills learned largely outside school involving low-income children are particularly challenging during the early stages of research and program development. Investment in the collection of mixed methods data may be critical for accomplishing the broader set of objectives outlined above. A project that has objectives of strengthening future measures, improving the effectiveness of programs, and in assessing the effects of alternate assumptions on evaluation outcomes needs mixed methods data to accomplish these objectives. Much of the long-term value of this project may still be ahead of us and lie not only with the "evaluation results," but in this future research with the data collected. Most of our mixed methods data has yet to be fully incorporated into the analysis and interpretation of results, and stronger recommendations for improving future research and the effectiveness of the WINGS program will emerge from continuing research with this data.

Our team of researchers originally viewed the value of this project as residing almost entirely in the evaluation results. However, we recognize that evaluations must broaden their scope to include better understanding the complexity of the lives of program participants, the complexity of managing and improving out-of-school programs to improve children's outcomes, the difficulty of measuring these early developing skills and the complexity of the analysis required to obtain and interpret results. The unique data collected in this project on the lives of low-income urban families, the development of their children in the K-1 period, and the impact of schools and programs on their lives can support years of research that can, (1) assess the sensitivity of current results to alternate assumptions, (2) assess the relationships among measures of early developing cognitive skills and academic outcomes, (3) improve current measures of early developing skills, (4) improve the WINGS program, and, (5) identify social and educational policies that can improve outcomes for low-income children.

Evaluations involving non-profit programs dependent on non-governmental funding require development of a strong researcher-clinician partnership. Directors of non-profit programs for children are in a challenging position because their outside funding from foundations and philanthropists often depends on successful evaluations. These sources of funding often do not have the sophistication required to understand the complexity of evaluating these programs and seek simple "black-white" answers (Did the evaluation show your program was effective at a $p < .05$ level?). Directors also often do not develop the requisite understanding of research to communicate effectively more nuanced results. But researchers also often have a "black-white" approach to evaluation and often use imprecise language in communicating results.

It is essential in these projects to develop the researcher-clinician partnership that allows both researchers and clinicians to better understand and interpret results. We came into the project with the traditional "hands-off" (let's do an "objective" evaluation, and not get involved in the program) approach to evaluation. We have determined partly through a conflictual process with WINGS and their able set of research advisors, that this "hands-off" approach was counter-productive and would not have allowed the project to proceed, nor would a fair evaluation have been possible.

Researchers have a lot to learn from clinicians, and vice-versa. We learned that results can be communicated in a more accurate way that better reflects the probabilistic outcomes of RCTs that program managers can also understand. We learned that program managers are very interested in results that reflect $p < .20$ and $p < .10$ because these results can point to ways of improving the programs, and may also be harbingers of future effects as measures strengthen and the power of samples increase. We learned that the mixed methods data collected can be used in many ways to offer suggestions for improving their program. We learned that, in our case, program managers can develop a pretty sophisticated understanding of research methods and offer suggestions and interpretations that strengthen the research. In particular, annual meetings of presentation of results allowed program managers and staff to suggest new, different, and sometimes better interpretations of the results than supplied by researchers. This ongoing sharing and interpretation of results at annual meetings was critical to building the partnership.

Chapter 9 Fidelity of Implementation and Program Improvement

9.1 Introduction

This chapter summarizes findings from the WINGS implementation and improvement study, which was designed to assess the context in which WINGS operated, the quality of program operations during the impact study data collection period, and to identify ways that the program could be improved. This chapter provides current preliminary results, but future research will be able to more fully incorporate all of the data collected to address the questions below.

These findings complement and supplement those of the WINGS impact evaluation. In order to interpret the findings of an impact evaluation, researchers need to put those findings into context. They need to understand how well a program is run and the contextual features that may contribute to its impact. If a program is well run, with adequate resources, and is appropriate for the population served, but program impacts are modest or non-existent, this suggests that the program design itself does not lead to desired outcomes. However, if a program is inadequately supported or run, or is unsuited to its participants, and it does not demonstrate impact, then determining whether it has the potential to bring about desired outcomes were it to be better run is difficult. Information from implementation studies that examine program operations can offer insights into the role operational quality plays in driving program impacts. The data collected on implementation combined with the results of the evaluation can also be focused on how to improve the program. For a program that offers promise that its current implementation is producing important benefits, the next step is to determine how to improve the program to both strengthen impacts on participants, to find ways of expanding and serving more children, and to establish a program evaluation for this expanded and strengthened program.

The WINGS implementation and improvement study's overarching questions are:

- What factors in the external context (school, community, and family) influenced the potential for program impacts?
- What factors in the WINGS organization influenced the potential for program impacts?
- How faithfully did program staff deliver the program to children?
- How could the program improve the impacts on children?

9.2 Data Sources

This chapter draws on several sources of fidelity information including:

<u>Enrollment and attendance information:</u> Information from the WINGS for Kids *Efforts to Outcomes* database was used to examine children's engagement in the WINGS program.

<u>Observations:</u> Each year, researchers observed the nests of children in the study approximately three times, using a variety of research tools. From 2012 through 2015, the Classroom Assessment Scoring System (CLASS, Pianta, La Paro, & Hamre, 2008) was used to assess the quality of emotional support and behavior management that each WINGS nest received. A tool developed specifically for measuring WINGS implementation fidelity was used to gather qualitative data during the 2015-2016 school year.

<u>Interviews and survey:</u> A survey of WINGS leaders was conducted in Spring 2016 and interviews with WINGS leaders, supervisors, and executive staff were conducted in Fall 2014 and Fpring 2016 to understand their experiences with the program.

9.3 School Changes

During the WINGS evaluation period, changes in every study school affected the implementation of the WINGS program. The changes include changes typical for schools, such as principal and teacher changes, but they also included more unusual changes, such as school relocations. These changes caused stress for families and school personnel and also affected children's enrollment and attendance in WINGS. At the same time that the schools experienced changes, the WINGS programs also changed. Again, there were typical changes, such as changes in WINGS leaders and program directors, but there were atypical changes, such as WINGS expansion to Atlanta, GA.

Appendix I shows a table that describes the changes that occurred at each of the schools between Fall 2012 and Spring 2015. Below we discuss why they occurred and their impact on the WINGS program. We then describe changes to the WINGS organization and their effects on WINGS implementation.

9.3.1 Two School Moves

During the second year of the project (2013-14) two schools moved locations from North Charleston to downtown Charleston. Prior to the study, both schools were located in downtown Charleston and had active WINGS programs, but the school district relocated the students and teachers for a few years beginning in 2011 in order to build schools that could more easily avoid damage from earthquakes. The initial move resulted in a greater proportion of North Charleston students attending both schools. Once the schools moved back to their original locations, the North Charleston students going to one of the schools faced transportation difficulties, especially getting home from the WINGS program. In the other school, the North Charleston students were not allowed to

enroll in it after the move since the school began enrolling only students from the local catchment area in Downtown Charleston.

9.3.2 School Pedagogical Changes

During the second year of the project (2013-14), one study school decided to adopt a Montessori model of instruction for preschool through 3rd grade. This created a significant shift in the demographic makeup of the student body, with many students being of higher socioeconomic status than those previously attending the school and in the study. This drastic shift in pedagogy also resulted in changes and turnover of teachers due to the required training of the Montessori method. The WINGS program for kindergarten was dissolved that year, and ultimately, WINGS ended its program at the school. At another school, a new principal hired in during the 2014-2015 school year began to change school standards in an effort to receive certification as an International Baccalaureate school. These schools were the rebuilt downtown Charleston schools.

9.3.3 Principal Changes

How much WINGS was or was not supported by schools varied from school to school, principal to principal, and year to year. During the second and third years of the project (2013-14, 2014-15) principal changes occurred in three schools; in two of those, the changes negatively impacted the implementation of WINGS. In one of the schools, the new principal appeared to be less interested in, and supportive of, WINGS, and when teachers complained that their classrooms were not left in the condition they wanted, she restricted the WINGS program to the gym and cafeteria. Due to the physical constraints, the schedule of WINGS activities (as well as the style with which they were implemented) changed dramatically. In the other school, the principal spearheaded the change to a Montessori school, which ultimately led to significant demographic shifts in the student population and the closure of the WINGS program at the school.

9.3.4 Schedule Changes

During the first year of the project (2012-2013), two study schools shared a building temporarily and as a result, one school was forced to change the start and end time of their school day, due to busing issues. This resulted in WINGS holding their program later in the day, and for a shorter period of time to avoid ending later in the day, which resulted in an overall lower program dosage for children in the school.

9.3.5 Changes Initiated by WINGS

Two important WINGS-initiated changes occurred during the study period. First, during the three-year study, WINGS expanded its program across three East Coast states. Impacts of expansion were felt during the first year of the study, when WINGS expanded to Atlanta, GA. Some program staff moved to Atlanta to take on leadership roles, requiring that additional program directors be promoted and trained in Charleston. The

summer training of WINGS leaders in Charleston was bigger, given that WINGS bused the trainees from Atlanta to Charleston to join the Charleston cohort in the training. Perhaps the largest effect of the expansion, however, was that key program supervisors and trainers spent a lot of time in Atlanta, which made it difficult for them to spend as much time with the Charleston staff as they had in the past.

Second, in the first year of the study (2012-13), WINGS informally changed its policy, which stipulated that program staff could ask children to leave the program for behavioral or attendance reasons. The new practice required that most of the children who would have previously been dropped from the program be retained to help minimize program attrition for the impact study. Overall, WINGS program staff did not like the change because they thought that it led to more behavioral problems in the program and that it allowed children who did not attend to take up space that other children might benefit from. In Year 2 (2013-14), WINGS returned to the old model of asking students to leave for behavioral and attendance reasons.

9.3.6 Changes Initiated by Schools

NORTH CHARLESTON ELEMENTARY SCHOOL (NCES)

School/ Administrative Factors: During the three year study period, NCES was stably located at the same site in North Charleston. NCES hired a new principal who remained until the spring of Year 3 (2014-2015) of the study, at which time, the assistant principal took over.

WINGS Factors: The principal was very open and welcoming to WINGS, and the WINGS program remained very solid during the three-year study period, with two strong program directors leading the WINGs program; the first for the two years, before the second took over in Year 3.

NCES was a very stable school with respect to location and administration and WINGS had a solid program during all three years

CHICORA ELEMENTARY SCHOOL

School/ Administrative Factors: During the three-year study period, Chicora was housed in the Robert McNair Building (3795 Spruill Ave). Chicora hired a new principal who remained in this role for the first two years of the study. During the third year the assistant principal moved into the role of Principal. She overhauled the teaching staff, bringing in many new teachers.

WINGS Factors: During all three study years the WINGS program had full support from both principals. Even though there was some turnover of Program Directors in Year 2, the program was very solid, extremely well run, and had a good reputation in the

community (families wanted to be part of it). The principal and PD had a great working relationship the entire time.

Chicora had a very stable school with respect to location and administration and WINGS had a very strong program during all three years.

MEMMINGER ELEMENTARY

School/ Administrative Factors: Several years prior to the study Memminger, originally located in downtown Charleston on Beaufain St., had moved to the Brentwood campus in North Charleston (2685 Leeds Ave.). At the Brentwood campus, Memminger shared a building with James Simons Elementary. This move was hard on families who lived downtown because the children now had to be bused to a new school. Additionally, Memminger enrolled all students living within its North Charleston vicinity, and thus mixed children from two different communities.

Prior to the start of the second year of the study (July, 2013), Memminger moved from the Brentwood campus in North Charleston back to its original space in downtown Charleston (20 Beaufain St), where it remained throughout the course of the study. This move was not completed by the time school began, with boxes still being unpacked on the first day of school. The school now did not accept children if they were no longer districted for the downtown location. As a result all North Charleston families previously attending were redistricted.

Memminger's principal remained for the first and second year of the study. During the third year of the study, a new principal was placed at the school. She began school procedures for International Baccalaureate certification. Some related standards made their way into school programs that year.

WINGS Factors: During the three-year study period there was a shift in program direction, with a new program director hired at the beginning of Year 3. The principal initially was fairly uninvolved with WINGS. However, she did express concerns during the end of Year 2 that property was being defaced in the classrooms. By the third year, she banned WINGS from using classrooms and only allowed use of the gym and the cafeteria. This forced WINGS to change its regular schedule in order to serve its many students within such rigid space constraints.

Memminger had external and internal disruptions that affected school staff and families from moving and transportation issues for families, administrative changes, and beginnings of pedagogical changes. WINGS also had to work around space restrictions and a schedule change (shortening of program)

JAMES SIMONS ELEMENTARY (JSE) SCHOOL

School/ Administrative Factors: Like Memminger (though smaller), JSE was originally a downtown school that moved from 741 King Street to the Brentwood Campus in North Charleston several years prior to the study. This move was hard on families who lived downtown (those originally districted to JSE), not only because the children now had to be bused to school, but the school had to stagger its start and end time with that of Memminger's. JSE families (who all lived downtown) complained about significant transportation issues, and their children getting home late and exhausted.

Midway through the second year, the school moved back downtown to its original address on King Street. The demographics of the school changed, with a greater percentage of the newly districted children being of a higher socioeconomic group. In addition, at the start of this school year, JSE switched to a Montessori model. All grades below fourth followed this model.

A new principal was hired during the summer before our study, and remained in her role for all three years of the study. In response to the shift to a Montessori model during Year 2, she made changes to her teaching staff, hiring new Montessori trained teachers and moving other untrained teachers around, either within or outside of the school.

WINGS Factors: During the first year (2012-13), JSE implemented a staggered start and end to the school day, due to sharing space with Memminger. The long school day that resulted forced WINGS to shorten its program to two hours. Families still complained about getting home late.

In Year 2, as a result of JSE's change to Montessori, WINGS did not hold a kindergarten group but still had first grade students. By the third year of the study, the WINGS program was no longer implemented at JSE. They had finished out the previous year and did not even start the next year because the demographics had changed so much.

JSE had schedule changes due to sharing schools in Years 1-2 that shortened WINGS and families struggled with children coming home late. In addition, the school shifted to a Montessori Model in Year 2 that changed student demographics, and WINGS was unable to offer a kindergarten program in Year 2, and completely closed the program in Year 3.

9.4 The Use of School Space

The expectations that school staff have for the use of their space are critical to the operations of after-school programs located in schools. The teachers and principal convey behavioral norms to students and after-school program staff, and those norms do not necessarily align with after-school program practices. School staff expects children to behave in an orderly and quiet fashion in the school hallways. Conversely, WINGS staff sometimes likes to encourage children to play in the hallways–for example,

encouraging them to flap their arms as if they are soaring birds—and those behaviors may not please teachers who remain in the building after school.

Also, when WINGS staff is overseeing Academic Center in the classrooms, school staff expects them to ensure that the children do not use school supplies intended for the school day or leave the classroom in disorder. The responsibility of ensuring that classrooms are left as they were found demands a large amount of WINGS staff's attention. The tensions between school and after-school staff regarding the use of space have been a longstanding issue in the use of school buildings for after-school programs (Walker et al., 2000), and there has been little change over the years. The WINGS staff does their best to respond to the expectations of school staff, and doing so affects program operations.

9.5 Program Fidelity

Program fidelity is a catch-all phrase that is generally used to gauge how much a social program's implementation conforms to the program model as outlined by its developers. It incorporates multiple dimensions, such as adherence, exposure, quality of delivery, participant responsiveness, and program differentiation (Dane & Schneider, 1998). Distinguishing these various dimensions is important in assessing the extent to which a program is run with fidelity. Confirming that staff delivered every activity in a manual does not provide sufficient information to conclude that a program operated with fidelity. If this was the case, but participants attended for only 25 percent of the time, then participants did not receive the full program. Here we briefly define each of the dimensions of fidelity.

Adherence is the extent to which activities are delivered as planned by the program developers.

Exposure refers to the number of sessions, the amount of time per session, and the duration (days, weeks, months or years) that the program lasts.

Quality pertains to how well a program is implemented. In after-school programs it often refers to the extent to which staff and children form good relationships, as well as the extent to which staff respond to children's needs in positive ways, include them in tasks, and encourage the development of new skills. Depending on the program's developers, program quality may be implicit or explicit.

Participant responsiveness is typically defined as how often participants attend the program.

Program differentiation refers to the extent to which a program is able to delineate how it differs from other programs that may have similar features. There are, for example, many different types of after-school programs, and many hope to develop

children's social-emotional skills. WINGS differentiates itself from those programs through its specific practices and strategies.

In our examination of fidelity, we address the following questions:

- Adherence: To what extent did WINGS leaders involved in the impact study adhere to the written WINGS curriculum and to their training? If staff modified activities, why did they do so?
- Quality: What was the overall quality of WINGS programming?
- Exposure: To what extent did the WINGS program provide the amount of programming that it planned to provide?
- Participant Responsiveness: What proportion of WINGS programming did children actually receive?

We will not include an extensive discussion of program differentiation because WINGS for kids is a well-differentiated program that includes a well-defined set of social-emotional skills divided into five major competencies (self-awareness, self-management, responsible decision-making, social awareness, and relationship skills), specific strategies for helping children develop those skills, and an organizational structure that has been designed to ensure that staff can carry out key tasks.

Before delving into how WINGS performed on each dimension, it is important to put forward some overall thoughts about program fidelity. Because fidelity encompasses so many dimensions, the overarching question "Was WINGS delivered with fidelity?" is difficult to answer. The answer to this question varies according to the dimension under consideration and across different WINGS activities and school locations. Nevertheless, the multiple answers provide rich opportunities for considering practical implications for the program.

9.5.1 Adherence and Quality

We discuss the overall quality of WINGS in conjunction with adherence to the model because adherence to the program model should correspond with program quality. Adherence to the WINGS model can only be understood in terms of what should ideally happen during a WINGS day. Appendix H describes what the program should look like.

To assess adherence to the program model, the research team used two instruments: an observational guide that focuses specifically on the extent to which staff used WINGS strategies to model social-emotional skills and guide children's development and the Classroom Assessment Scoring System (CLASS). During the 2015-2016 school year, all kindergarten and first grade WINGS nests in the remaining two study schools were observed three times using both the CLASS and the observational guide along with two nests of older children (grade 3 and grade 5). In total, nine WINGS leaders and four peace managers were observed.

9.5.2 The Quality of Staff-Child Interactions

In this section we discuss indicators of program quality based on observations of practices recorded using the CLASS tool. The CLASS measures interactions between the staff member and the children in a group, focusing primarily on the teacher's behavior. It incorporates three major domains of a positive classroom environment: emotional support, classroom organization, and instructional support. Within each are three or four dimensions that capture key aspects of the overarching domain. Instead of using all domains and questions in the CLASS, we have selected those that represent skills that WINGS program managers reported they explicitly addressed in their WINGS leader program manual, provided training to site staff on, and coached WINGS leaders on. We omitted all items related to instructional support. Teaching does happen at WINGS, in particular with respect to social emotional skills. However, to date, program managers have not focused greatly on how to train WINGS leaders in instructional support, nor is such training used as a basis for hiring WINGS leaders. In fact, students who are working on education degrees are less likely than other students to become WINGS leaders because their classroom and teaching practicum schedules conflict with the WINGS program schedule. Researchers scored the observations in accordance with the pre-k CLASS manual.

We must exercise caution in assessing the WINGS program using the CLASS. The tool was initially developed to observe classroom teachers, not after-school programs. While there are many elements, such as positive climate, that are desirable in both arenas, the CLASS is oriented toward classroom teaching. Even in areas of seeming overlap, the alignment between desired features of an after-school program and the CLASS may not be strong.

Table 9.1 provides average scores (on a scale of 1-7) for each selected item by year. A high score (i.e., 6-7) on all items but one is considered good. The one exception is negative climate, for which a low score (i.e., 1-2) is good. For all items, scores between 3 and 5 are considered mid-range.

Table 9 . 1 : CLASS scores across years

	2012-2013	2013-2014	2014-2015
Positive Climate (PS)	4.4	4.5	4.2
Negative Climate (NC)[1]	1.8	1.7	1.5
Teacher Sensitivity (TS)	4.7	4.0	4.0
Behavioral Management (BM)	4.4	4.2	4.4
Productivity (PD)	4.8	4.2	4.7
Instructional Learning Formats (IFM)	3.8	3.5	3.8

Scores are in the mid-range on most items, with the exception of negative climate, in which scores are very low, which is desirable[1]. These low scores correspond to qualitative observations that WINGS leaders tended to be calm, even when children acted out.

Scores for instructional learning formats (which is part of behavioral management, even though it addresses the teacher's facility with a variety of instructional skills) tended to be worse than scores on other items (although still in the mid-range of quality). There is relatively little change from year-to-year, although negative climate dropped even lower over time, indicating very few instances of irritability, sarcasm, or disrespect by teachers or children in the program and no major instances of negative interactions. Teacher sensitivity also dropped over the course of three years, which was not desirable.

No definitive reasons seemed to account for the changes observed from year to year with respect to program quality. There was more change at the school level than across schools, but no patterns emerge from the qualitative analysis that might explain those changes. Among the changes that we hypothesized might make a difference in terms of how the program was implemented were changes in two schools' location, which resulted in changes in the schools' population. We also hypothesized that turnover in WINGS program directors at the schools might influence how well the program ran, but no consistent patterns emerged.

We observed modest differences in scores across nest ages (see Table 9.2). Looking at differences that are at least .5 (on a 7-point scale), we see that teacher sensitivity and

---

[1] Overall, negativity scores on the CLASS tend to be very low and have little across studies which suggests that it is a poor measure. However, in WINGS, those low scores were supported by qualitative observations that described WINGS leaders as very calm, which is discussed in more detail below.

positive climate tend to be higher in kindergarten-only classrooms. In particular, teacher sensitivity and positive climate tended to be higher in all activities for nests that included only kindergarten children than for all other age groupings. For all other items, differences in scores were .3 or less.

Table 9 . 2 Class Scores by Grades of Children in Nests

Grade(s) of Children in Nests	Positive Climate	Negative Climate	Teacher Sensitivity	Behavior Management	Productivity	Instructional Learning Formats
Kindergarten only	4.7	1.6	4.6	4.3	4.5	3.8
Kindergarten plus first grade	3.6	1.4	3.6	4.4	4.8	3.6
First grade only	4.2	1.7	3.8	4.2	4.6	3.6

Scores varied somewhat across activities (see Table 9.3). Scores on negative climate were remarkably similar across activities. In contrast, scores for instructional learning formats tended to be lower for Academic Center than for other activities. Lower scores in Academic Center probably reflect the fact that, instead of facilitating students' engagement in lessons and actively teaching, WINGS leaders are expected to facilitate students' homework completion. Behavior management scores were slightly higher in Academic Center than other activities, providing evidence of WINGS leaders' emphasis on keeping children in their seats and away from the regular school-day teacher's classroom materials, rather than on instruction.

Table 9 . 3 . Average CLASS Scores Across Activities for Selected Measures

Activity	Positive Climate	Negative Climate	Teacher Sensitivity	Behavior Management	Productivity	Instructional Learning Formats
Academic Center	4.2	1.6	4.2	4.5	4.7	3.3
Choice Time	4.6	1.6	4.1	4.2	4.3	4.0
Discussion	4.3	1.7	3.9	4.0	4.3	3.9
WINGS Works	4.5	1.6	4.2	4.2	4.6	3.9

9.5.3 The Level of Adherence to the Program Model

While adherence to the program model encompasses many activities, WINGS executive staff emphasized a few aspects as critical to the program. There needs to be a transitional period from the school day to after-school, Community Unity, which needs to be well run, and–because more talking is encouraged in WINGS than during the school day–program staff needs to change children's frame of reference for behavior and emotional climate. Lessons and discussions about social-emotional learning skills must be delivered as planned. All staff should reinforce social-emotional learning skills over time and in multiple activities. Across all activities, staff interactions with children need to be warm and caring. Staff should also manage children's behavior in positive ways. Finally, while not necessarily central to WINGS' social-emotional model, Academic Center is critical in that it is essential for the program's acceptance by parents and teachers, and because it takes up a substantial amount of children's time each day.

In this section we emphasize these critical components as we consider the degree to which programs achieved adherence. This section addresses adherence to the model through qualitative data collection on WINGS-specific topics. Where appropriate, we combine that information with information from the CLASS.

Is Community Unity well-delivered?

The tone set in Community Unity on any given day can make or break the overall quality of the day's program. The presence and skills in working with children of all staff are key to the success of the activity, although the program director's skills are arguably the most important. The program director and assistant must be able to command attention and generate enthusiasm and positive energy among the children. The WINGS leader must be present, actively monitoring interactions among children and making eye contact to encourage participation. If staff members are absent or not actively engaging with the children, children's behavior deteriorates.

Observations are consistent in noting that the program directors and assistants regularly demonstrated enthusiasm and energy and kept children engaged in activities. They danced, played music, commented on children's positive behaviors, corrected children's (quite rare) negative behaviors in appropriate ways, and generally managed the children in the room very well. This pattern is unsurprising given that program directors and some program assistants had previously been WINGS leaders and had several years' experience with children. Program directors, in particular, are hired due to their excellent skills in managing groups of children.

On the other hand, WINGS leaders' activities in Community Unity varied considerably with respect to quality, probably due to their lack of experience relative to other staff. WINGS leaders often had no experience managing large groups of children before they joined the program. At times, they struggled to translate the games that program

directors introduced into ones that younger children could play successfully. At other times, the WINGS leaders could not keep children's attention on the activities. There was, however, only one observation when the WINGS leader lost control of his group during Community Unity. In general, when observers noted that there were problems in Community Unity, they were related to the skills with which program assistants or WINGS leaders explained games to the entire group. While program directors typically emceed the session, occasionally they delegated the work to someone with less experience, which led to fewer instances of positive reinforcement and to unclear explanations of activities. This was a likely driver of the relatively low CLASS scores for instructional learning formats. Even under those circumstances, the sessions were otherwise well run and upbeat. Program directors remained present in the room and managed children, albeit in a quieter way.

Are lessons and discussions about social-emotional learning skills delivered as planned?

Typically, one of the 30 social-emotional learning objectives in the WINGS manual is introduced each Monday in Community Unity. Discussion is also held on Mondays; it offers an opportunity to go into more detail and ensure that children have an understanding of the week's objective.

Overall, the WINGS leaders followed the Discussion lesson plans in the manual and engaged children in discussions about their experiences. For kindergarten and first-grade children, doing so meant reading a book that highlighted the weekly objective and asking questions about the book. Once the book was finished, the WINGS leader discussed the children's experiences with them or played a game that illustrated the concept.

The primary challenge to fostering children's social-emotional development was the challenge WINGS leaders faced in reinforcing lessons throughout the course of the year. The WINGS model calls for WINGS leaders and other staff to remind children about skills and use teachable moments to reinforce the practice of those skills in the weeks and months after they are introduced. Those moments can be affirming when children use the skill appropriately or correcting when children need to be reminded to use the skill in particular situations. However, WINGS leaders, who manage groups of around 10-12 children, do not have time reinforce all the social-emotional learning skills in the course of day-to-day activities. Instead, they tend to pick a subset to highlight. Skills that they emphasized regularly throughout the year included the following:

- Focus on activities;
- Manage and understand your impulsive behavior;
- Take responsibility for your actions instead of blaming others; and
- Consider the perspectives of others.

Overall, WINGS leaders were much more likely to emphasize three of WINGS' five major domains of social-emotional learning: responsible decision-making, social relationships, and self-management. In an analysis of qualitative notes from observations of 2015-2016 WINGS nests, WINGS leaders mentioned objectives in those dimensions 39-51 times each, whereas objectives in self- and social-awareness were mentioned 10-11 times each.

It seems from these observations that WINGS leaders tend to reinforce the socio-emotional learning lessons that matter when managing groups. This may seem necessary in order to maintain control of nests during activities. While helping a child understand his or her emotional state can help the child identify alternative behaviors and solve conflicts that arise, we speculate that in the group settings the WINGS leaders did not have the time to work with children individually.

Peace managers, who are responsible for helping children manage their behavior when WINGS leader cannot, have more time to interact individually with children and engage with them about awareness of themselves and others. However, our observations suggest that they often do not use all of the tools at their disposal to do so. Of the four peace managers observed, all asked children why they engaged in their behavior, but only one consistently asked how else they could have handled themselves, discussed the options available to them, and then encouraged them to go back to their nests and try different behaviors next time. The remaining three told the children how they should behave differently in their nests.

In addition, WINGS staff often model positive feedback by thanking children for their good behavior and being very explicit about what the children did to receive praise.

Is Academic Center well managed and productive?

Academic Center is a key element of daily WINGS activities. Teachers, parents, and principals all value the inclusion of homework completion time in the program. Children go to Academic Center four days a week for fifty minutes, so it comprises a significant portion of WINGS program time.

The program model calls for children to work alone on their homework. Children are to be encouraged to read books when done with their assignments. WINGS leaders provide homework help.

As noted earlier, Academic Center tends to receive relatively low scores on the CLASS observations on instructional learning formats and higher on behavior management compared with other activities. Its scores on other qualities are similar to those of other activities.

Responses to a survey of 31 WINGS leaders conducted in Spring 2016 offer insight about what might be contributing to these low scores. The activity poses challenges for the WINGS leaders. About one-third of respondents identified Academic Center as the most stressful part of their day in WINGS, which was twice as many as those who identified any other part of the WINGS day as most stressful. This is notable since WINGS leaders generally find their job stressful: 30 percent reported that they found their job stressful most or all of the time, and about 50 percent reported that they found the job stressful sometimes during a typical week.

Interviews with WINGS leaders who worked with kindergarten through second grade children provide some reasons why their job, and Academic Center in particular, can be so stressful. First, WINGS leaders identified Academic Center as one of the most frustrating activities for children. They reported that many WINGS children were behind grade level in math and reading, so they needed substantial support to complete their homework. WINGS leaders' did not always have the capacity to offer help to everyone who needed it during each session.

Second, WINGS leaders needed to engage in more behavior management during Academic Center than other activities for several reasons, further contributing to children's frustration. When children needed help with their homework and the WINGS leader was helping another child, children sat at desks or tables with little to do but raise their hands and chat with other students. Under these circumstances, it was not unusual for one or several children in the kindergarten or first grade nests to get up from their seats and wander to talk to other students, look at classroom displays, or ask to go to the bathroom. Because Academic Center is supposed to be a quiet time when children work on their own, the WINGS leaders responded to these behaviors by directing children to go back to their seats. Also, if children finished their homework ahead of time, then the WINGS leaders encouraged them to read books to themselves, but there were typically children in the room who did not want to do so. This resulted in additional off-task behaviors that the WINGS leaders had to manage.

Finally, when Academic Center was held in regular classrooms, as it typically was, WINGS leaders needed to be very careful to leave the classroom as the school day teacher had left it. But many of the classrooms had age-appropriate toys for children, and it could be challenging for the WINGS leaders to keep the children away from them. All of these events triggered the need to employ behavior management strategies, which took time that WINGS leaders could have used to work with students on their homework or interact in other ways.

In support of the possibility that holding Academic Center in teacher classrooms caused stress for WINGS leaders, WINGS leaders at Chicora reported that Academic Center was less stressful than any other activity. This is the one school in which the children were not allowed in the teacher classrooms after the principal received numerous

complaints from teachers about how their classrooms were left. In response, she had WINGS conduct Academic Center in the auditorium.

This alternative setting presented both drawbacks and advantages. WINGS leaders often had the children sitting in single rows of ten, and the children didn't have desks. Also, the noise level in the room was often high because of the number of children. On the other hand, being in one room meant that WINGS leaders could support each other more easily. They could call on the peace manager or program director for help with child behavior. Our data are limited, but of the five WINGS leaders we interviewed, only one preferred the classroom to the auditorium, despite the auditorium's disadvantages.

How effectively do WINGS leaders manage behavior?

Several characteristics of how WINGS leaders worked with children stand out. Observers commented across multiple observations that WINGS leaders were calm when they responded to children, which aligns with the low negative climate scores on the CLASS. Nine WINGS leaders were observed two to three times each in the 2015-2016 school year, and observers noted that six of those leaders were calm and quiet when responding to children who were not doing what the WINGS leader had asked or what was expected. In the remaining three nests, the observers did not observe that the leader was agitated, but did not remark on how calm they were. Commonly, WINGS leaders used eye contact and reminded children what they needed to do (e.g., "control yourself"). Less commonly, WINGS leaders diverted children's attention by asking them to do a task or answer a question.

Another strategy that WINGS leaders commonly used was to disregard mildly disruptive behavior. The central office WINGS staff that oversaw refinements to the program model and trained WINGS leaders reported that they had emphasized in training the importance of disregarding behavior that did not cause harm to the child or others, since it took time away from the leaders' capacity to pay attention to children who were engaging in activities and behaving well.

What was children's possible exposure to WINGS?

In any program intervention, exposure to the program includes both the planned schedule and the actual schedule.

- **Planned schedule**: For how many hours a day is the program planned? For how many days a week? For how many weeks? And for how many years?
- **Actual schedule**: Planned schedules do not always consider unexpected events that may disrupt program activities. What disruptions occurred in WINGS over the course of the study?

WINGS' planned schedule is five days a week, for three hours a day over the course of the entire school year. In practice, while the school district is in session for 180 days,

WINGS is open approximately 165 days per year, and does not meet when schools are in session for half days. Thus, a full year of WINGS is about 500 hours.

Over the course of the evaluation, several major disruptions occurred that upset WINGS schedules in the schools. Perhaps most importantly, two schools were co-located while their original buildings were rebuilt. Given space limitations, one school's WINGS program was shortened by an hour per day, meaning that children received only two-thirds of the planned WINGS dose in that school in one year. In that same school, the kindergarten WINGS program was closed mid-year due to changes in the school's curriculum. Kindergarten children who were in WINGS during the 2013-2014 school year and attending that school could therefore have received only about one-third of the planned program hours during the second year, while the first graders received about two-thirds. This may have negatively affected the results of the impact study across schools.

9.6 Summary of Findings

- The need to meet requirements of school-day staff in using school space posed significant challenges to achieving program fidelity. Important challenges included the need to maintain orderly classroom spaces and conflicts between behavioral expectations during the school day and during WINGS.
- Program components generally adhered to the program model. Major deviations typically involved modifications of activities to improve suitability for kindergarten students.
- Quality of implementation varied across dimensions, activities, and school locations. Programs performed especially well in the area of low negative climate. Programs tended to be less strong with regard to instructional skills of staff.
- Quality of delivery tended to be relatively low for Academic Center, reflecting the fact that it was a somewhat problematic activity for both WINGS leaders and children. A common challenge was the need to balance behavior management in the classrooms with homework help.
- Quality of delivery of a key program activity, Community Unity, was adversely affected when the relatively inexperienced WINGS leaders took the lead.
- Probably due to the need to maintain a controlled group atmosphere, WINGS leaders focused much more on domains of social-emotional learning that are important for managing groups (responsible decision-making, social relationships, and self-management) than on areas related to encouraging introspection (self- and social-awareness).
- WINGS leaders displayed many effective behavior management practices, including calmly responding to children and disregarding mildly disruptive behavior.
- WINGS' actual schedule was at points in the study period significantly reduced from the planned schedule. This meant that some children received as little as one- to two-thirds of planned exposure to the program in certain years.

9.7 Preliminary Implications

9.7.1 Implications for Program Evaluation

The evidence presented from the implementation and improvement analysis suggests three areas that have implications for the analysis of results and their interpretation. These areas are: school -evel differences, differences between cohorts and differential staff focus among pro-social outcomes.

The implementation analysis suggested that the quality of the program and the capacity of the children and parents to benefit from the program was different across schools and changed over time. In particular, two schools (Memminger and JSE) were relocated during the study that triggered busing of some children long distances, changes in the composition of the students and parents, variation in program quality, and reduced dosage levels. JSE also changed to a Montessori model that dramatically shifted the demographics of their students, and WINGS ended their program in the third year of the study. On the other hand, NCES and Chicora had more stable environments, facilities, and student characteristics over the entire course of the study. This suggests a hypothesis that should be tested that Chicora and NCES should have stronger effects than Memminger and JSE.

The implementation analysis also suggests that there were more disruptions that could have impacted the quality of the WINGS program for Cohort 1 than Cohort 2. The school relocations impacted Cohort 1 more than Cohort 2. In addition, the WINGS program expanded to Atlanta in the study's first year and program leaders and staff spent much time in Atlanta and relocated key staff from Charleston to Atlanta. Finally, a key policy on asking children/parents to leave the program changed over the course of the study. Prior to the initiation of the study, WINGS strictly enforced attendance and behavior policy that resulted in asking children/parents to leave the program. When the study started, this policy was relaxed, but later during the second year it reverted to the stricter policy. This shift was partly in response to WINGS staff who thought a stricter policy led to fewer behavior problems. If so, Cohort 1 would have more behavior issues that might affect program quality than Cohort 2. The combined effects of school relocations, WINGS expansion and WINGS Policies would all favor Cohort 2 having stronger effects than cohort 1. This hypothesis can also be tested with our data.

The third area where the implementation analysis contributes to the analysis and interpretation is the suggestion that differential staff focus was on three of the five domains of socio-emotional learning. The three domains that appear to receive emphasis were responsible decision-making, social relationships, and self-management, while the two that received less emphasis were social and self-awareness. This suggests the hypothesis that stronger effects should be present for the three domains receiving emphasis.

9.8 Implications for WINGS

Consider putting the youngest children (kindergarten and first graders) in cafeterias and/or auditoriums for Academic Center. The use of classrooms for Academic Center is challenging for young children and WINGS leaders. The children are distracted by the classroom materials, and the WINGS leaders must focus on keeping the children from touching anything. Having the younger children in areas where there are fewer distractions may very well decrease stress for the WINGS leaders.

Develop a more defined WINGS leader career pattern. Turnover among WINGS leaders is very high from year to year, and developing a career pattern for leaders may be helpful in addressing turnover. Developing specific competencies, acknowledging and celebrating WINGS leaders' growth in those competencies, and having more levels of WINGS leaders with associated wage increases (e.g., Wings Leader I, Wings Leader II), may encourage year-to-year retention. There will be costs associated with this, but those costs would be much lower than hiring professionals.

Provide more and different activities for different ages of children.
Currently, the WINGS curriculum provides largely the same activities for all age groups, with some minor modifications. This presents several challenges to WINGS leaders. First, given their inexperience in working with children, the WINGS leaders cannot easily modify activities so they are more age appropriate. When the children can't understand the activity it becomes hard for the WINGS leaders to engage with them. Second, children are doing the same activities year after year. Given the complexity of the program and the wide range of ages, it seems to make sense to codify age appropriate modifications and activities in the manual and curriculum, instead of leaving those changes up to the WINGS leaders. While a skilled classroom teacher with experience across several age groups could probably modify activities, the WINGS leaders do not. Codifying the changes in the manual and curriculum would likely make the WINGS leaders' jobs easier and reduce their stress. That, in turn, might foster more positive interactions with children and increase retention from year to year.

Pair WINGS leaders on a regular basis. WINGS leaders occasionally work in pairs, and interviews suggest that they generally enjoy the support they get from partners. Working in pairs suggests that WINGS leaders have their nests in close proximity so that they can engage in shared activities. This would allow leaders to share their strengths, which will likely help them engage and manage the children. The range of skills required of a strong WINGS leader is fairly broad—s/he must be a kind and calm adult presence whose skills include behavior management, problem solving, performance, and child engagement, academic support, and sensitivity to child dynamics. Careful pairing of WINGS leaders can help ensure that as many skills as

possible are available to children and help WINGS leaders develop skills in areas they are not strong in initially.

Pairing and training/coaching practices would need to be carefully considered. It would make more sense to pair experienced and skilled WINGS leaders with less experienced leaders to allow the more experienced leader to model for the less experienced leader. These dyads are likely to need help forging useful partnerships, and thus program directors would need to be trained in coaching the teams.